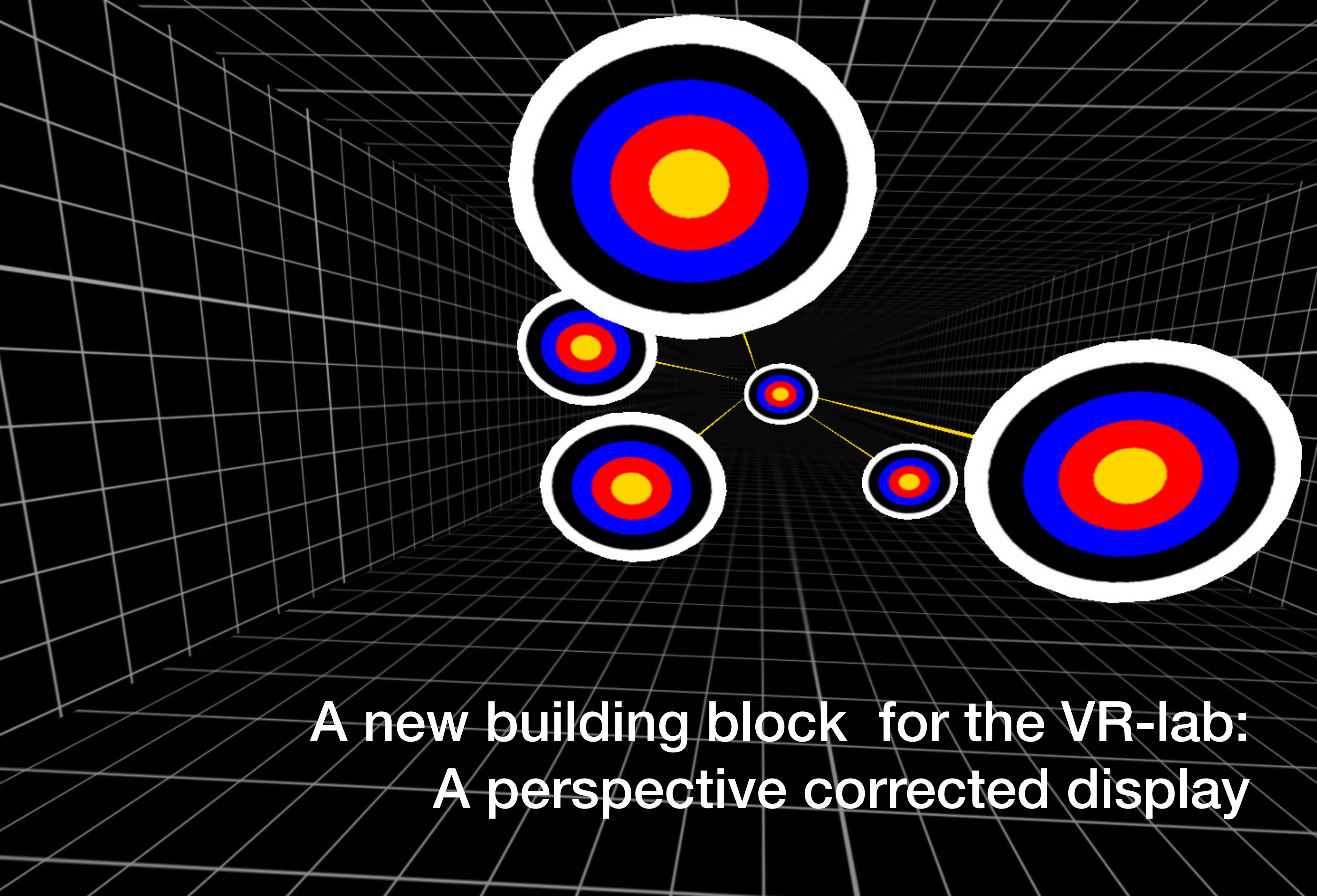
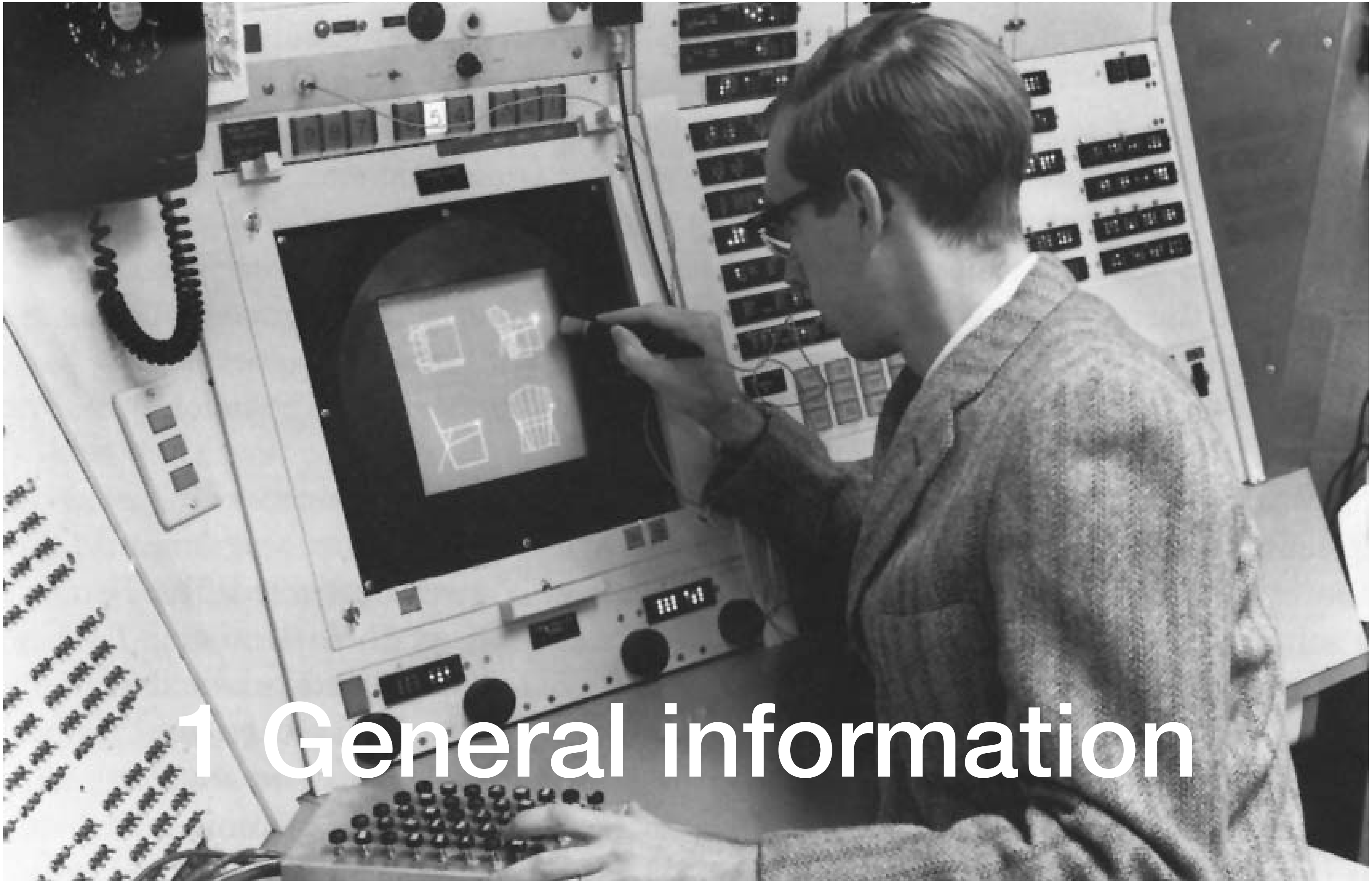


A new building block for the VR-lab: A perspective corrected display



1 General information

1.1 Report information

This report contains the findings and results of my Bachelor Thesis. The assignment is to design and make a perspective corrected display system for the Virtual Reality lab of the University of Twente.

This report is addressed to:
Damgrave, msc R.G.J.
Boer, prof.dr.ir. A. de

Date:
04-11-2011

Number of pages:
106

Number of appendices:
3

Executing student:
Jurriën Dijkstra, s0088013

figure 1.1: Ivan Sutherland's Sketchpad was a milestone in real-time graphics

1.2 Introduction

The following assignment is a bit different from your average bachelor thesis. It is customary within the bachelor industrial design to execute the thesis within a company. Often a (consumer) product is designed and the design process one has learned is followed which ultimately leads to a prototype of the product.

This thesis is executed within the University of Twente. An opportunity opened up to execute an assignment within its virtual reality laboratory. I felt this was a unique chance, since it is not very likely to find an environment like this outside the University. I have great affinity with electronics and over the years I acquired quite some knowledge on this topic, so I had no doubts an assignment would fit like a glove.

There were no readily available assignments, so ultimately I could propose what I wanted to do and the VR-lab would or would not agree. It did not take long to find my ultimate use of virtual reality within the field of industrial design:

As designers we often make use of digital technologies; especially CAD models are very helpful. What I still miss is the ability to walk around an object and see it in its natural context. I feel it would be fantastic to see a virtual object on a true scale, independent of my location using augmented reality. By making use of a retinal image display and smart sensors the model should always appear in a correct perspective to the user and the environment. It could also have a consumer use: imagine walking in your virtual kitchen, while your house is still unfinished.

Unfortunately, this use of VR will not see the day of light within this thesis; the VR-lab is principally against using glasses. What the VR-lab proposed instead, is to develop a perspective corrected display, like the one popularized by Johnny Lee with the Wiimote. I agreed, since in essence the assignment is still the same, only the outcome is different.

This thesis will follow a process analogous to a regular design process, although the focus is a bit different. Instead of desk research, a lot of scientific papers have been delved through and an assessment tool, instead of sketches, leads to a design. The final outcome is a software environment as opposed to a tangible object. This thesis has proved to be a difficult one, not in the least because of stubborn hardware and software. But eventually things have come together and the virtual reality lab has now gained another building block! (Albeit in need of some more elaboration)

I hope this report will enlighten the reader on the topic of perspective corrected displays and tracking technologies. It is meant to give insight in all the variables that play a role in the design of a perspective corrected display and present the reader with an overview.

Jurriën Dijkstra, 2011

figure 1.2: Spatially augmented reality

1.3 Table of contents

1.4 Summary

This report has been written within the framework of the bachelor thesis of the study industrial design at the University of Twente. The assignment has been executed for the virtual reality lab of this university and comprises the design of a “perspective corrected display”. In such a system three-dimensional images, of which the perspective corresponds with the position of the user, are shown on a legacy two-dimensional screen.

First of all chapter two shortly addresses the history of virtual reality to get an idea of the framework the system will be used in. When the foundations of VR are outlined, a more specific look will be taken at the different elements that play a role in this assignment. A “perspective corrected display” is also sometimes referred to as a geometric display, because it uses projective geometry to create a three-dimensional illusion. Several geometric displays are discussed to find out which facets play a role in the design thereof. This category of displays is available in many shapes and sizes, but is currently mostly used within the academic world. The element that is the most important in the realization of a perspective corrected display is the method which is used to track the user.

Since the tracking method is such an important element of the design it is looked at extensively in the third chapter. The focus is on the inner workings of a technology to clarify why some variables have an influence on the performance of the system. All tracking methods discussed in this chapter are capable of tracking a user’s head position accurately and with low latency, but not on all accounts. There appear to be many variables that play a role which can influence the performance of a tracker tremendously.

In chapter four it is tried to give insight into a large number of those variables. Applications, screens, the type of user, the environment and motion trackers are reviewed. This time, the focus is mainly on the consequences a variable can have on the system (both positively and negatively). All parameters congregate in an assessment tool, in which the applicable ones can be filled in step by step. Depending on the choices that have been made this tool shows what problems could arise. Based on this information and on previously acquired knowledge the best solution route for the system can be chosen.

The best solution route is not always feasible within a certain timeframe and budget. That is why in chapter five some readily available tracking methods and software environments are shortly tested. After testing it becomes apparent that the Wii-mote suffers from compatibility issues and is almost unusable. Tracking based on a webcam delivers some promising results, but needs improvement.

Chapter six can be seen as the equivalent of a detailed design in a regular design process. A pragmatic look is taken at how the development of a geometric screen can be realized. The fully functional software FaceApi is used for the tracking part of the system. Next, the acquired data is send via an UDP-stream over the network and converted to a stream that complies with the TCP protocol. This is a necessity, because the VR-software Quest 3D can only listen to streams that are based on the TCP protocol. The imported data is modified in such a way that it leads to camera movements that correspond with the position of the user.

To serve as a fully-fledged building block in the VR-lab, the system designed in this thesis should be further extended. The basic functionality is present and the acquired effect is convincing, but the solution route could be more efficient. The commercial version of FaceApi offers more functionality and eases the sending of data. The modifications that are performed in Quest 3D can be wrapped in a so called “channel” which enables easy integration in other VR-projects.

1.5 Dutch summary

Dit verslag is geschreven in het kader van de bachelor opdracht van de opleiding industrieel ontwerpen op de Universiteit Twente. De opdracht is uitgevoerd voor het virtual reality lab van deze universiteit en behelst het ontwerpen van een "perspective corrected display". In een dergelijk systeem wordt op een standaard tweedimensionaal scherm een driedimensionaal beeld getoond waarvan het perspectief overeenkomt met de positie van de gebruiker.

Allereerst wordt er in hoofdstuk twee kort ingegaan op de geschiedenis van virtual reality om een idee te krijgen van het kader waarin het systeem zal worden gebruikt. Wanneer de fundamenten van VR omlijnd zijn wordt specifieker gekeken naar de onderdelen die in deze opdracht een rol spelen. Een "perspective corrected display" wordt ook wel een geometrisch scherm genoemd omdat de afgebeelde geometrie aangepast wordt om de 3D illusie te creëren. Verscheidene geometrische schermen komen aan bod om erachter te komen welke facetten een rol spelen bij het ontwerp daarvan. Deze categorie schermen is er in vele soorten en maten, maar wordt vooralsnog voornamelijk gebruikt in de academische wereld. Het onderdeel dat de belangrijkste rol speelt in het realiseren van een perspective corrected display is de methode waarmee een gebruiker gevolgd wordt.

Aangezien de tracking methode zo'n belangrijk onderdeel is wordt hier uitgebreid naar gekeken in het derde hoofdstuk. Er wordt gefocust op de werking van een techniek zodat duidelijk wordt waarom sommige variabelen een invloed hebben op de prestatie van het systeem. Alle behandelde tracking methodes zijn in staat accuraat en zonder vertraging het hoofd van een gebruiker te volgen, maar niet in alle gevallen. Er blijken veel variabelen een rol te spelen die de prestaties van een tracker in grote mate kunnen beïnvloeden.

In hoofdstuk vier wordt getracht een groot aantal van die variabelen inzichtelijk te maken. Applicaties, schermen, het type gebruiker, de omgeving en bewegings trackers passeren de revue. Dit maal wordt vooral gekeken naar de gevolgen die een variabele kan hebben voor het systeem (zowel in positieve als negatieve zin). Alle parameters komen samen in een beoordelingstool, waar stapsgewijs ingevuld kan worden welke van deze aanwezig zijn. In de tool wordt afhankelijk van de keuze getoond welke problemen zich voor zouden kunnen doen. Op basis van deze informatie en eerder opgedane kennis kan de beste oplossingsroute worden gekozen voor het systeem.

De beste oplossingsroute is niet altijd haalbaar binnen een bepaald tijdsbestek en budget en daarom wordt in hoofdstuk vijf kort getest met enkele al beschikbare tracking methodes en softwareomgevingen. Na testen blijkt dat de Wii-mote compatibiliteitsproblemen ondervindt en VRijwel onbruikbaar is. Tracken met de webcam levert op zich belovende resultaten, maar moet nog wel verder verbeterd worden.

Hoofdstuk zes kan worden gezien als het equivalent van een detailontwerp in een normaal ontwerpproces. Er wordt op pragmatische wijze gekeken hoe het geometrische scherm tot stand kan komen.

De kant en klare software “FaceApi” verzorgt het tracking gedeelte. Vervolgens wordt de verkregen data via een UDP-stream over het netwerk gestuurd en omgezet naar een stream volgens het TCP protocol. Dit is noodzakelijk omdat het VR-programma Quest 3D alleen naar streams op basis van TCP kan luisteren. De geïmporteerde data wordt zodanig verwerkt dat dit leidt tot camerabewegingen die overeenkomen met de positie van de gebruiker.

Om als volwaardige bouwsteen in het VR-lab te voldoen, zal het hier ontworpen systeem verder uitgewerkt moeten worden. De basisfunctionaliteit is aanwezig en het verkregen effect is overtuigend, maar de oplossingsroute kan efficiënter. De commerciële versie van het programma FaceApi biedt meer functionaliteit en vergemakkelijkt het verzenden van data. De bewerkingen in Quest 3D kunnen verpakt worden in een zogenaamde “channel” waardoor het systeem eenvoudig in andere VR-projecten kan worden geïmplementeerd.

2 Virtual reality

2.1 Assignment

Virtual environments have become a commodity and they are increasingly realistic. The VR-lab features a variety of virtual environments that enables users to analyze and test products and/or situations. Most virtual environments can be controlled by user input, ranging from a conventional keyboard to multitouch interfaces. The downside of using input devices that need to be operated by the user is that they can distract the user from the actual task. Ideally, a VR application should be so intuitive that it can be used without prior training and be totally invisible to the users, allowing concentration on the task, and not in using the tool.

One way to make control of an application easier is to use direct manipulation. The application should exhibit the following characteristics: Objects of interest must be continuously represented, physical actions are used instead of complex commands, and actions should be immediately visible. Another research suggests making the interaction with the virtual environment as analogous to real life experiences as possible. *(Davies R C, 2004)*

To realize these design guidelines, one has to look beyond conventional input devices that require the user to actively control the device. This type of input device would disconnect the user from the technology and the user should preferably be unaware of the device. The user should be able to navigate in an environment or evaluate an object by just moving around. Meaning, the VR application will always offer a correct view of an environment or an object based on the position of the user.

The goal of the assignment is to design and make a new building block for the Virtual Reality lab of the University of Twente, which supports realistic control of virtual environments.

2.2 Introduction to VR

Virtual Reality is a comprehensive term which describes the technology and the field of application in general. It enables users to interact with virtual environments in real time. Virtual environments are artificial environments that are created with software and presented to the user in such a way that the user suspends belief and accepts it as a real environment. The more immersive a virtual environment is, the easier it is accepted as a real environment. Sensory cues provided by the virtual environment can improve immersion. For example, a head movement results in a different view in the VE and touching objects in the VE results in haptic feedback. Input devices, ranging from keyboards to motion trackers, can be used to navigate in and interact with the VE. (*How Stuff Works,2011), (Wikipedia, 2011), (Cybertherapy, 1998)*

One of the earliest examples of VR, in the modern meaning of the word (i.e. involving electronics), was developed by cinematographer Morton Heilig. He wanted the future theater to stimulate all the senses to make the viewer feel like he was in the movie instead of watching a movie. He detailed his vision in a paper in 1955 and in 1962 he developed a prototype of his vision called the Sensorama. This VR device included a stereoscopic display, fans, odor emitters, stereo speakers and a moving chair. Users would take place in this one-person theater and experience one of five two-minute 3D movies. Real-time graphics where unavailable at the time, thus everything was prerecorded. One of the films showed on the Sensorama was a motorcycle ride through Brooklyn in the 1950s where the wind blows through your hair, you hear the city sounds, you smell the city and you feel every bump in the road. The Sensorama was not successful however, since it proved to be difficult to develop a solid business model for this kind of machine. (*Carlson W, 2008), (Wikipedia, 2011)*

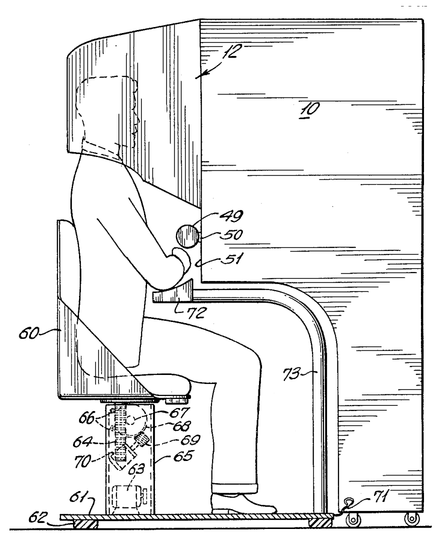


figure 2.1: Sensorama patent drawing

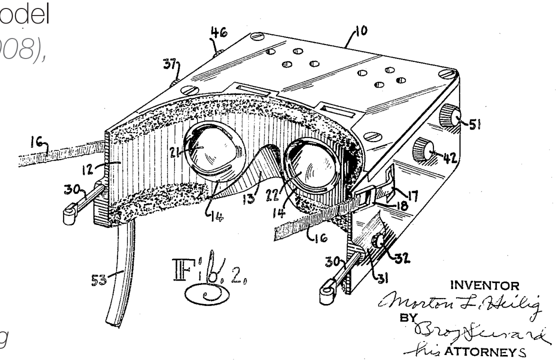


figure 2.2: Telesphere mask patent drawing

figure 2.3: Sensorama brochure

In 1960 Heilig also proposed an idea for a head mounted device with a wide field of view, stereo sound and an odor emitter. This device, called the "Telesphere mask", was the first patented head mounted display (HMD), but it was not the first one that was fabricated. In 1961 Comeau and Bryan, employees of the Philco Corporation, constructed a head-mounted display that they called the "Headsight". This device was developed to remotely view dangerous situations and showed that virtual reality could also be used for other purposes than entertainment. The system used magnetic tracking to measure the head direction of the user and showed the corresponding video images on a single CRT screen mounted on a helmet.

(Haller et al, 2007)

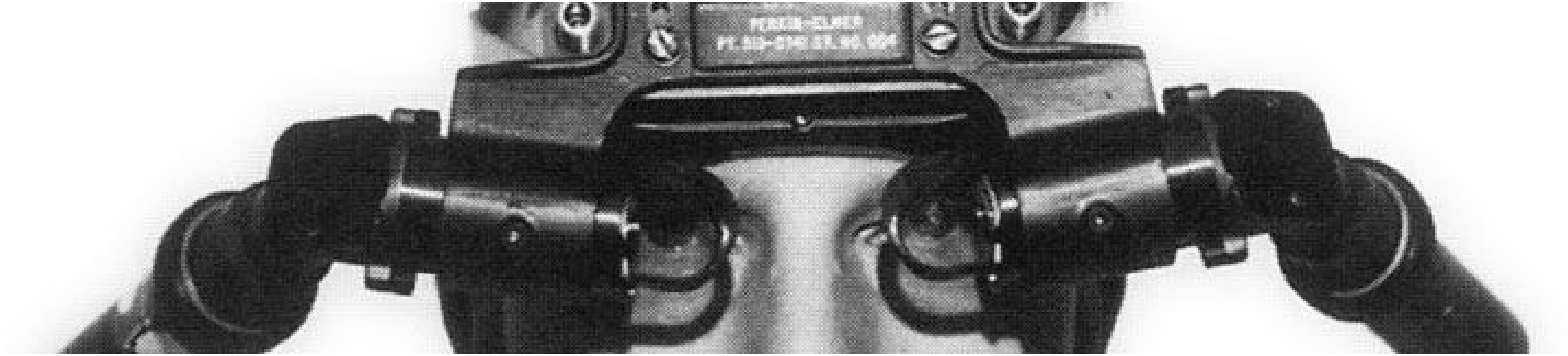


figure 2.4: Ultimate display

In 1965, computer scientist Ivan Sutherland envisioned the first scientific VR device in his paper the "Ultimate display". He predicted that advances in computer science would eventually make it possible to engineer virtual experiences that were convincing to the senses.

"A display connected to a digital computer gives us a chance to gain familiarity with concepts not realizable in the physical world. It is a looking glass into a mathematical wonderland. With appropriate programming such a display could literally be the Wonderland into which Alice walked".

(Sutherland I E, 1965)

He brought some of his ideas in practice only a few years later, when he created what is widely considered to be the first virtual reality and augmented reality head mounted display system. The "Sword of Damocles" got its name because it was too bulky to head mount. Instead, it was suspended from the ceiling and a user's head could be strapped to it. It tracked the user's head movements and provided stereoscopic imagery on two see-through CRT displays - one for each eye. The major breakthrough in this HMD was the use of a computer to generate the graphics in real-time.

This device laid the foundations for most future developments in VR:

- A virtual world that appears real to any observer, augmented through three-dimensional sound and tactile stimuli.
- A computer that maintains the world model in real time.
- The ability for users to manipulate virtual objects in a realistic, intuitive way.

Although many of the breakthroughs in the early history of virtual reality involve head mounted displays, this thesis will focus on desktop VR, i.e. involving a (legacy) screen.

(Carlson W, 2008), (How Stuff Works, 2011)

2.3 Benefits and limitations of VR

figure 2.5: Illustration of VIEW research

According to VIEW, a three year long project that was funded by the European Union, the use of virtual reality is potentially beneficial in the following situations:

• VR may be used where manipulation of real-life variables would risk damage to people, equipment or the environment (eg. in hazardous environments)

• VR can provide an alternative where manipulation of real-life variables has an unacceptably high associated resource cost (eg. logistics, finance, personnel or national security)

• VR allows rapid prototyping and configuration of an environment where perceptual input might need to be changed frequently (eg. reconfiguring interface details or instrument layouts)

• VR may be used to enhance or degrade or otherwise alter some aspect of reality. As such VR might be used to impose visual restrictions on the user (eg. smoke effects in a burning compartment or reduced visibility due to a visual defect), or VR might be used to highlight components in the real world which could otherwise be missed (eg. fire extinguishers or escape routes through buildings)

• VR allows users to experience views of micro or macroscopic entities in a variety of dimensions (eg. using VR to explore the atomic surface of materials or human tissue at a nanoscopic level (Stone, 2001b), in a way that would not be physically or ethically possible in the real world)

• VR can be used where real locations are impossible or difficult for users to physically occupy (eg. hazardous environments, simulating outer space, under the sea, etc)

(VIEW, 2003)

2.4 Volumetric or geometric?

There are two paths that can be chosen that lead to a device that satisfies the guidelines in chapter 2.1; volumetric displays and geometric displays.

A volumetric display device is a graphical display device that forms a visual representation of an object in three physical dimensions, as opposed to the planar image of traditional screens that simulate depth through a number of different visual effects. Because an object or environment is represented in three dimensions, multiple users can walk around it and have a very realistic experience. As defined in a scientific paper, "A volumetric display device permits the generation, absorption, or scattering of visible radiation from a set of localized and specified regions within a physical volume." Most volumetric displays are auto stereoscopic; that is, they produce imagery that appears three-dimensional without the use of additional eyewear. The volumetric analogue to pixels is called voxels, short for volume elements, or volume pixels. *(Favalora et al, 2001)* Volumetric displays provide perceptually rich 3D information by satisfying all visual depth cues; however they are challenging to implement, and current technologies are limited in resolution, brightness, opaqueness, and/or compactness.

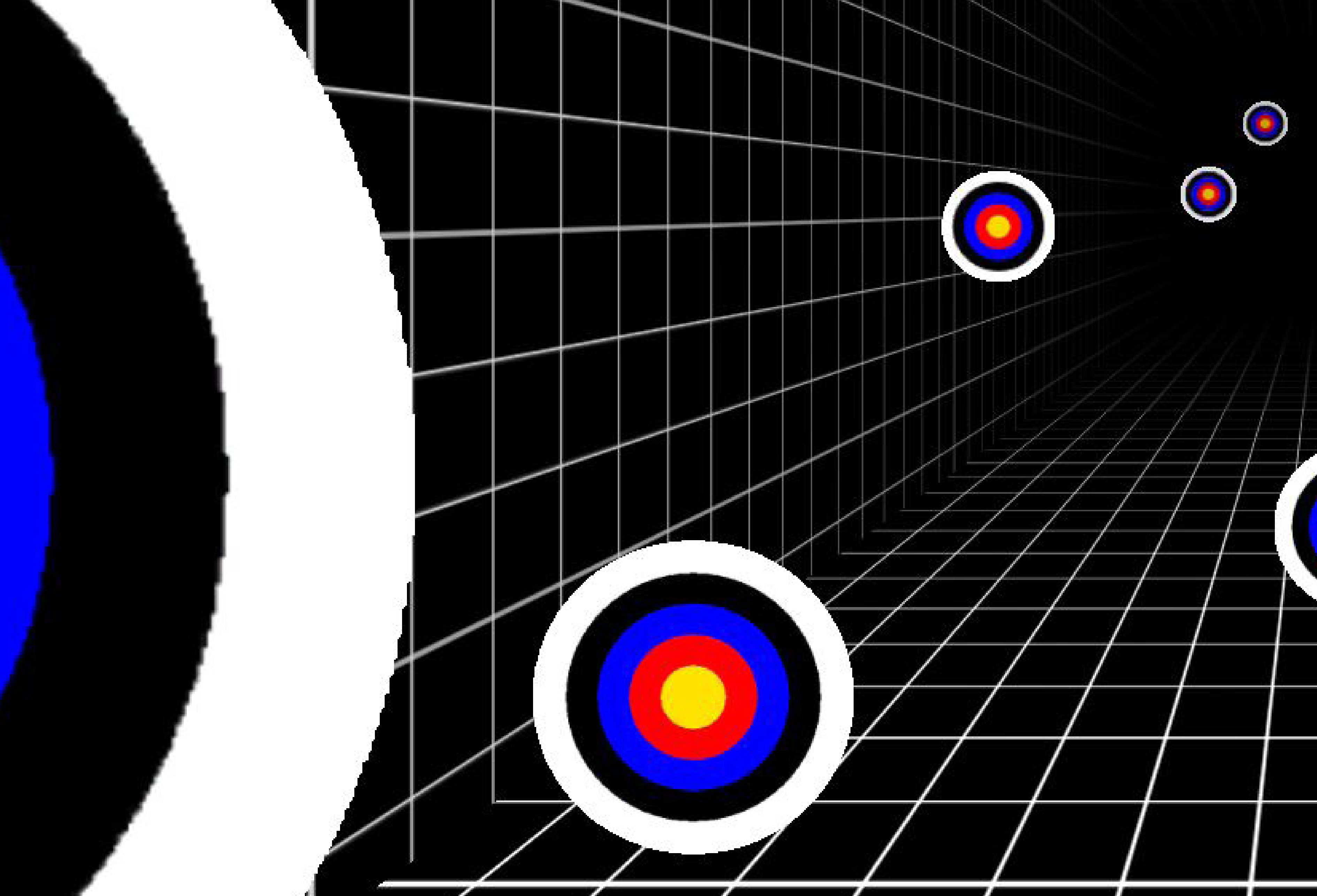
figure 2.6: A volumetric "persistence of vision" globe

A different approach is to render an image on one or more (legacy) 2D displays with a perspective corrected to the user's view. These displays can be classified as "geometric" as they use projective geometry to establish the illusion of 3D on a 2D surface by satisfying one or more perceptual depth cues. The imagery on a 2D screen changes in accordance to the position of a user in front of that screen. That is why a geometric display is also referred to as a perspective corrected display.

(Stavness et al, 2010)

Volumetric displays are very comprehensive and free a user of technology, but are not very suitable to use as a building block. Not all VR applications would fare well on current volumetric displays and the possibilities to adapt the device to a certain application are limited. Besides, the VR-lab of the University of Twente is already experimenting with a volumetric device. This thesis will take the path of the geometric display. Although it could be difficult to provide as much perceptually rich information as volumetric displays, geometric displays offer a cheaper alternative with great flexibility. This type of technology can potentially turn every screen in the VR-lab into a perspective corrected display.

figure 2.7: A geometric building projection



2.5 Geometric displays

The Virtual Reality Lab of the University of Twente will experiment with volumetric displays in the near future. A different approach to 3D that this research will focus on is to render an image on one or more 2D displays with a perspective corrected to the user's view. These displays can be classified as "geometric" as they use projective geometry to establish the illusion of 3D on a 2D surface by satisfying one or more perceptual depth cues.

The imagery on a 2D screen changes in accordance to the position of a user in front of that screen. That is why a geometric display is also referred to as a perspective corrected display.

2.5.1 pCubee

pCubee is a prototype that uses the Fish Tank Virtual Reality concept. FTVR uses head-coupled perspective rendering, stereoscopic techniques, or both, to provide optical cues to improve users' perception of 3D virtual environments. Stavness, Lam and Fels tried to extend one-screen Fish Tank VR by using a multiscreen setup arranged in a box shape.

Correcting the perspective of each screen to the user's head position gives the illusion of having real 3D objects within the box. pCubee allows a user to interact with dynamic virtual scenes that react to display movement with simulated physics in real-time. As a user manipulates, shakes and rotates the display box, objects within the scenes slide and bounce around.

pCubee has a wired magnetic tracking system (Polhemus Fastrak) to achieve low latency tracking and prevent lag. Lag has shown to disrupt the three dimensional effect. The head tracking sensor is embedded in the top of a set of headphones and the user's eye position from the sensor is estimated with a fixed offset.

In a user study subjects suggested a variety of application possibilities dealing with 3D visualization such as 3D radar, gaming, maps, storytelling and education.

(Stavness et al, 2010)

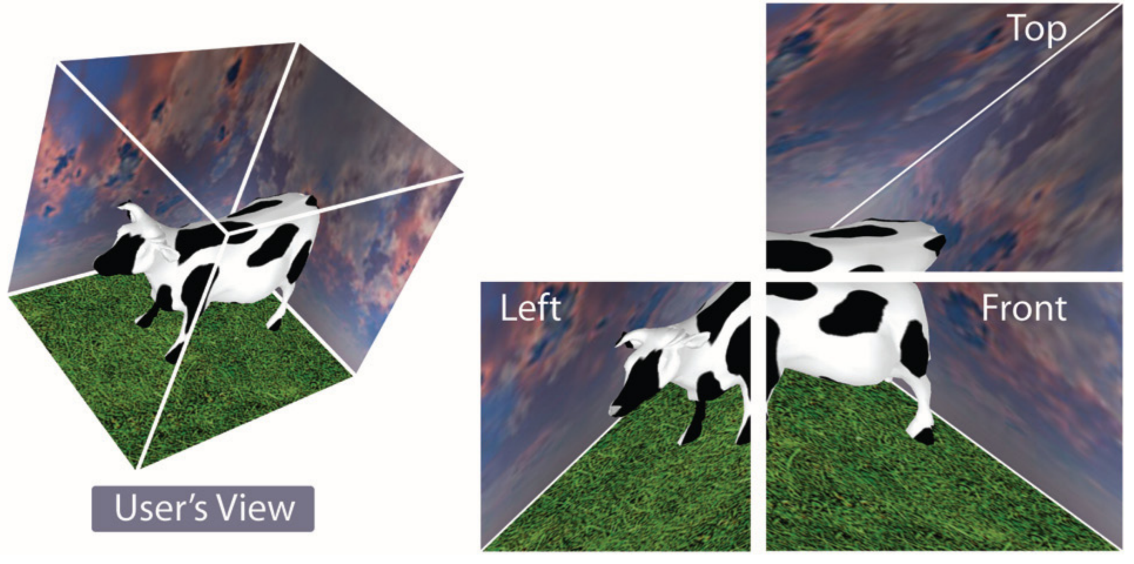


figure 2.8: How pCubee works

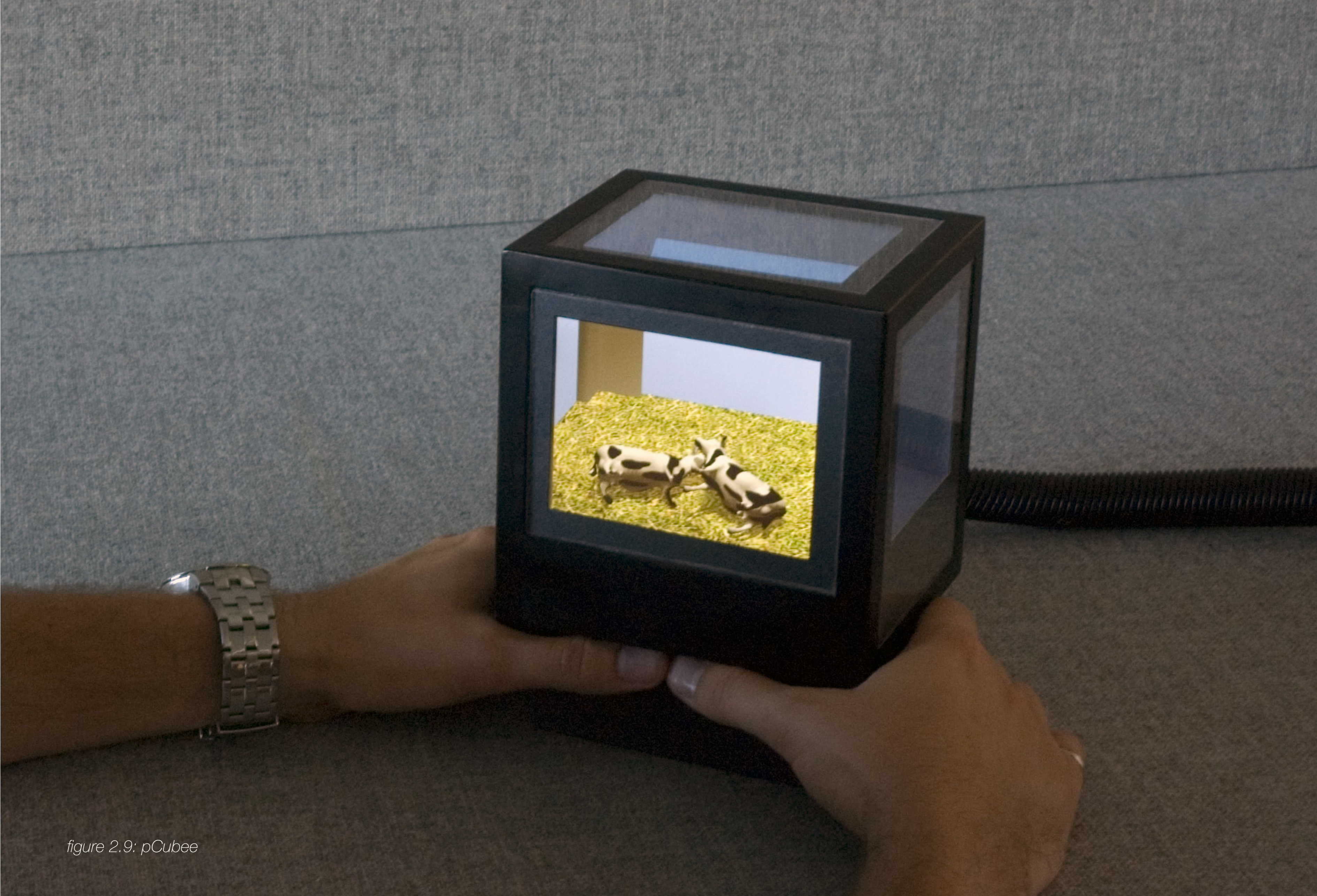


figure 2.9: pCubee

2.5.2 gCubik

gCubik is a handheld cubic display that achieves an autostereoscopic effect with integral image rendering and a lens array overlaying the screens. The motivation for this research was that collaborative tasks that require the sharing of an object can benefit from a compact, group-shared autostereoscopic display.

gCubik provides correct position-dependent multiuser viewing, while it does not require special glasses or head tracking for viewing. The display consists of three 3.5-inch LCD panels with VGA resolution, each with an IP (integral photography) micro lens array at the top.

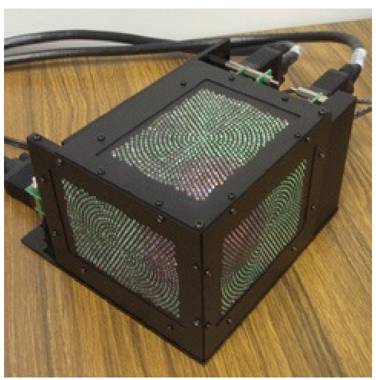
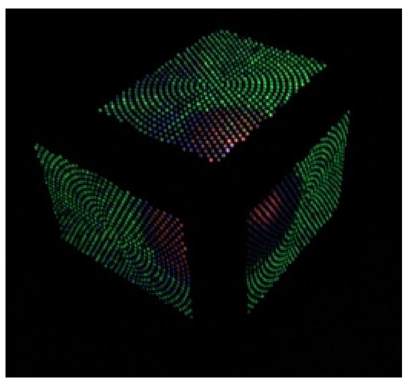
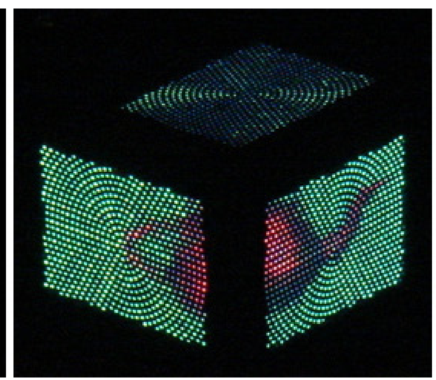
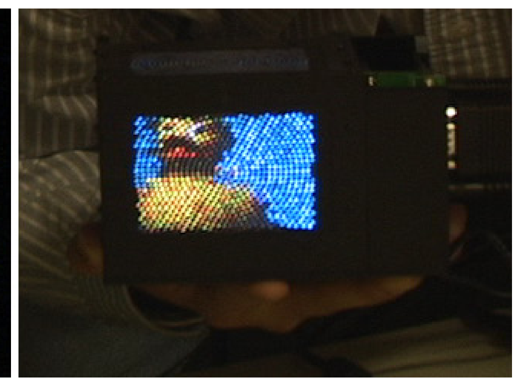
Each lens looks different depending on viewing angle. Thus rather than displaying a 2D image that looks the same from every direction, it reproduces a 4D light field, creating stereo images that exhibit parallax when the viewer moves.

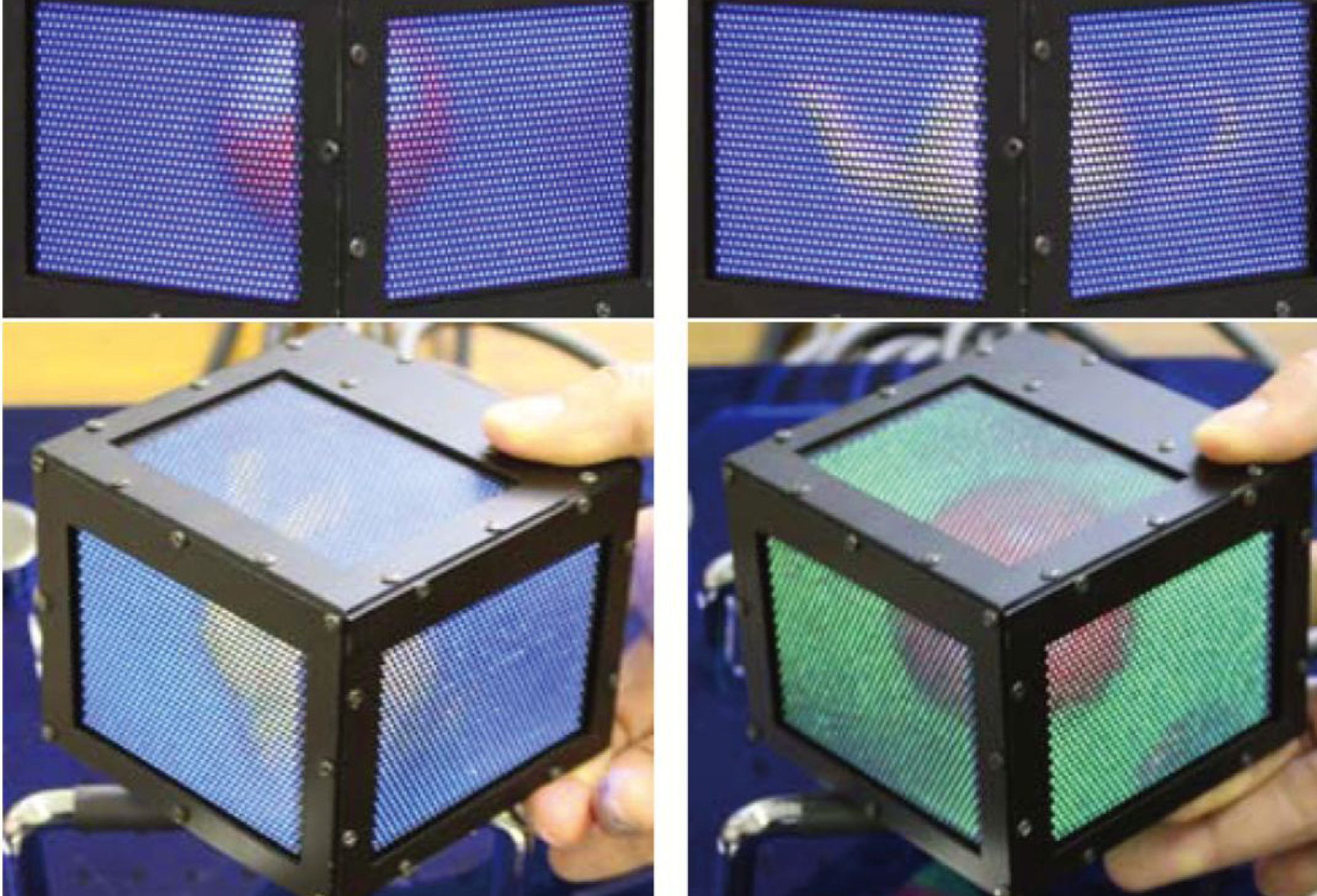
The researchers envision games and edutainment to be application areas.

(Lopez-Gulliver et al, 2008)

figure 2.10: gCubik

figure 2.11: gCubik



2.5.3 Cubby

Cubby is a desktop virtual reality system that features head-coupled motion parallax on three orthogonal screens. In contrast to pCubee or gCubik, where the objects are shown within the cubes , the objects that are displayed on Cubby seem to stand in between the screens. The three-dimensional illusion in Cubby is based on motion parallax and pictorial cues such as shading (Phong), occlusion and texture.

According to the Dutch researchers that developed Cubby one of the problems with motion parallax is that virtual objects move outside the display area and compromise the 3D effect. That is why they chose to use two vertical screens and one horizontal screen for Cubby.

Cubby uses back projection screens to prevent occlusion of the images. A Dynasight infra-red tracker was used to track head movement. The device tracks the position of a reflective target with a diameter of 6 mm that can be attached directly to the head or to a pair of glasses.

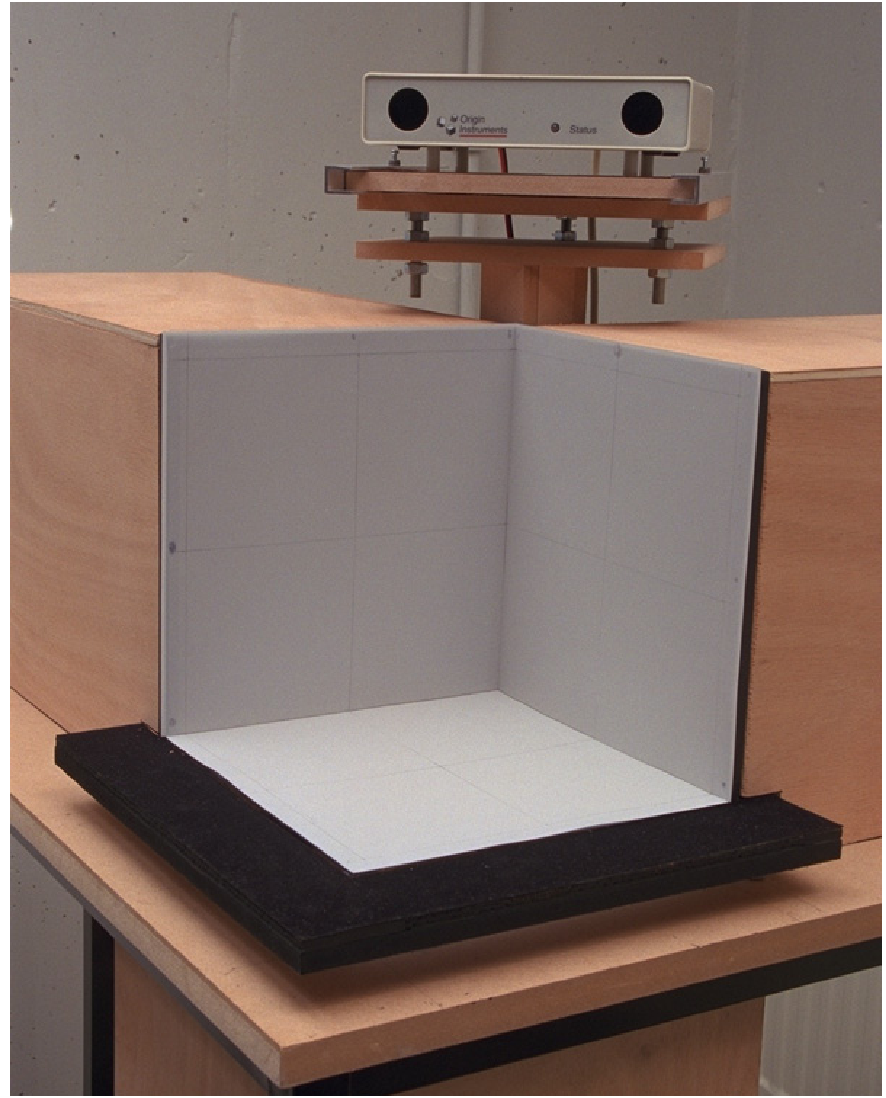


figure 2.12: Cubby

Because the virtual objects hover in front of the screen it is possible to implement direct manipulation of an object. In contrast to manipulation of virtual objects through an input device such as a mouse or a space navigator (3d connexion), proprioceptive information obtained in Cubby while manipulating virtual objects with an instrument can resemble the proprioceptive information obtained during instrumental manipulation of objects in everyday life.

(Djajadiningrat et al, 1997)

figure 2.13: How Cubby works

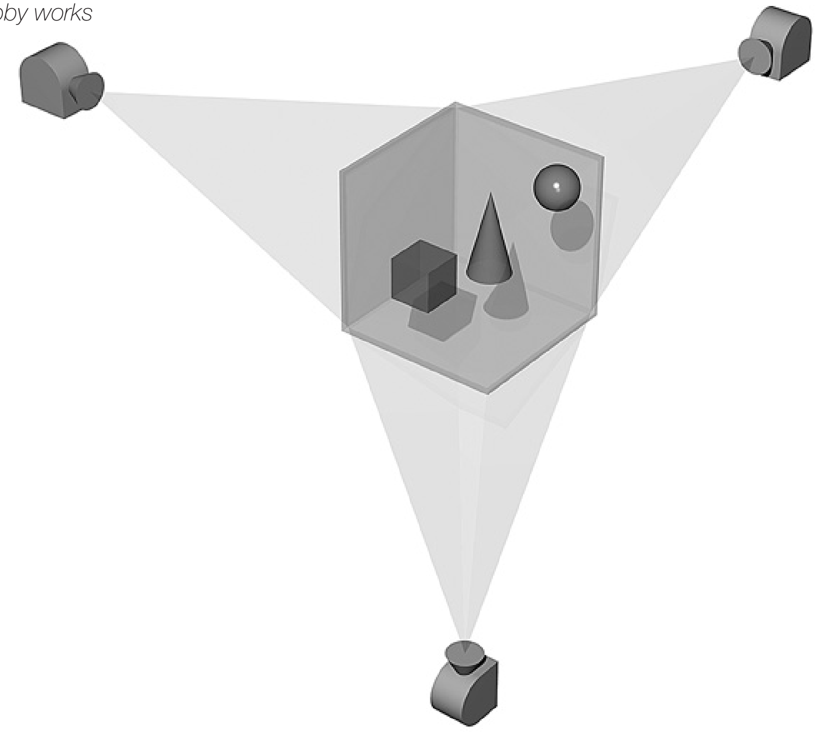
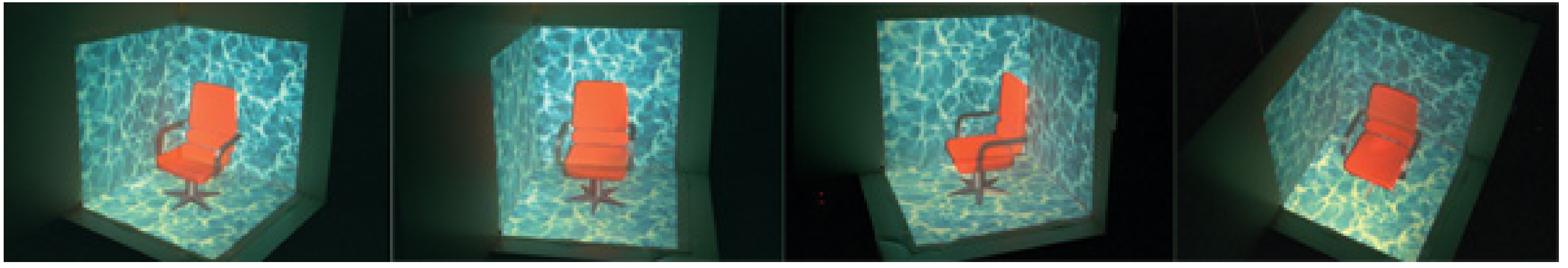


figure 2.14: The perspective is corrected to the user's position



2.5.4 Cave

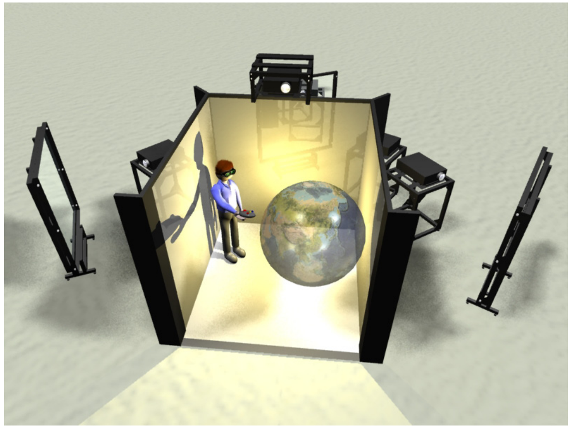
The CAVE (CAVE Automatic Virtual Environment) was designed to be a useful tool for scientific visualization. The CAVE is a 3m x 3m x 3m box shaped theater which is made up of three rear projection screens for walls and a down projection screen for the floor. The user stands inside the box and experiences immersive virtual reality.

The users head and hand are tracked with Polhemus or Ascension tethered electromagnetic sensors. Besides motion parallax, CAVE uses binocular disparity as a depth cue. Sound is also available, although at the time CAVE was built, they could not yet integrate directional sound (which could also have served as a cue).

The electromagnetic sensors are low latency and not very encumbering, but there is an important downside to this type of sensors. The electromagnetic trackers required building the CAVE screen support structure out of non-magnetic Stainless steel (which is also relatively non-conductive). But non linearity's are still a problem, partially because conductive metal exists on the mirrors and in the floor under the concrete. Wheelchairs, especially electric ones, increase tracker noise and non-linearity's as well.

A video showed that the experience proved to be very immersive; a user who was standing on the level floor inside the CAVE almost fell down because of the convincing imagery. It also shows that the CAVE can be used for other VR appliances than scientific visualization. *(Cruz-Neira et al, 1993)*

figure 2.15: How Cave works



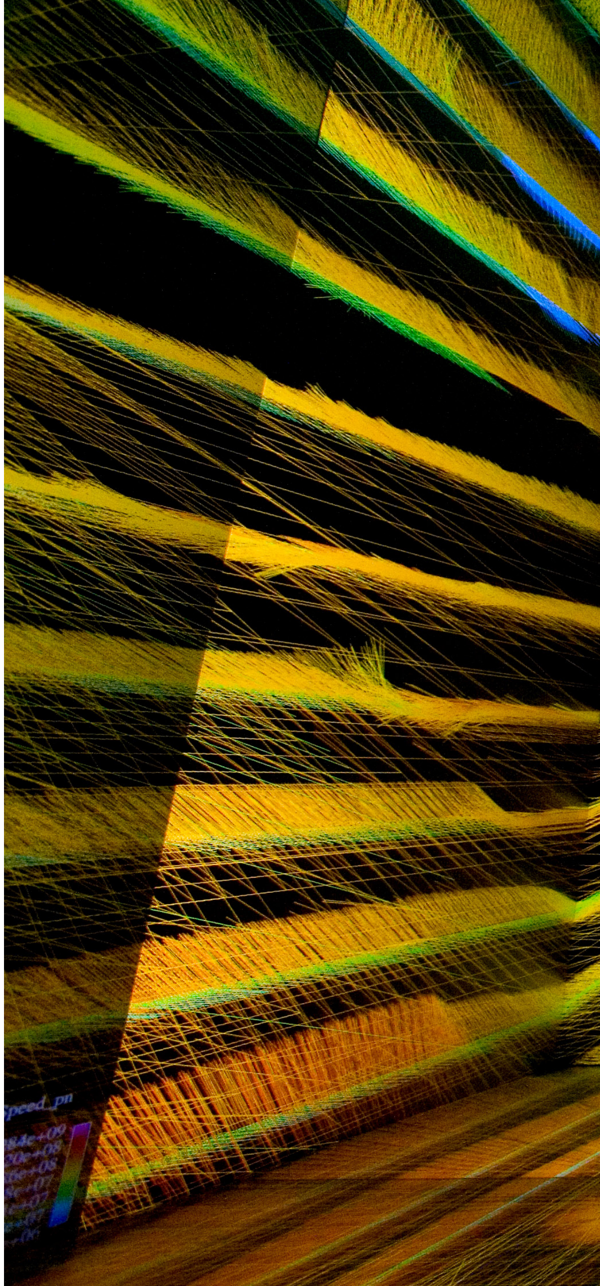
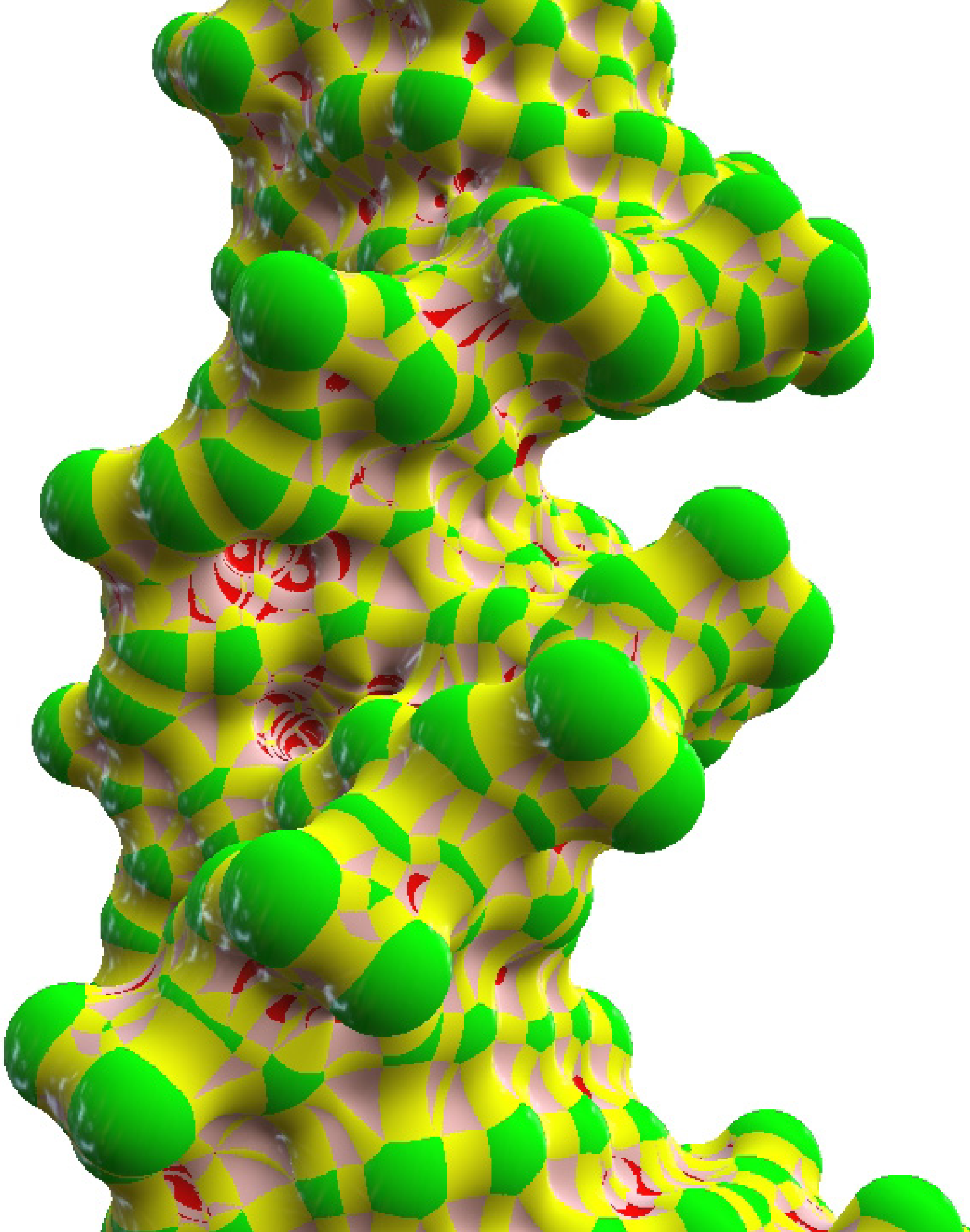


figure 2.16: inside CAVE

2.6 Applications

The application area of geometric displays is quite broad. Most beneficiary are applications where intuitive interaction and/or perceptually rich information are important. Research has been done to support this theory, but no closing evidence has been found. E.g. in one paper performance suffers when 2D depth cues are removed, which includes motion parallax. Some pages further in the article it concludes that motion parallax has no influence at all. *(Luo et al, 2007)*

To find some useful applications of perspective corrected displays a brainstorm session was held. The results of the brainstorm session can be found in Annex B.



2.7 Conclusions

Geometric displays can have various appearances and are very flexible in their use; a geometric display can be a tiny handheld device that supports group use, or a very large cube that engulfs a single user. For a geometric display to work as a perspective corrected display it needs to know the user's head or eye position. There are several tracking possibilities that each have their up- and downsides. The demands the tracking system has to fulfill depend on the application and the environment, but a general demand is that it should be a low latency system. A common depth cue used in PCD's is motion parallax, but it can also be combined with other depth cues such as binocular disparity.

Although the best type of screen to use depends on the application it is not responsible for the actual user input. The most important aspect of geometric displays and in particular perspective corrected displays is the tracking system. What tracking solution is the best depends not only on the characteristics a tracking technology exhibits, but also on the environment the application is used in. The key to designing a system which supports realistic control of virtual environments is to map all variables that influence the system.

First it is important to get a better understanding of the tracking possibilities and characteristics, which will be discussed in chapter 3. Chapter 4 elaborates on the variables that play a role in the VR-lab which results in a tool that assesses the best type of tracking depending on what variables are present.

figure 2.17: A scientific visualisation

3 Tracking technologies

3.1 Introduction

In this chapter several tracking methods are reviewed, using the information found on specific geometric displays as a starting point. The focus is on the technology in general, not on the large amount of variations that are present within a technology.

The different paragraphs within this chapter try to answer one main question: How does it work? Answers to that question have been found, however they might require prior knowledge of the field. Sometimes reading a paper or patent about a certain technology raised even more questions, especially in regards to the underlying principles of a technology. By understanding the foundations of a tracking method, it becomes clear why some variables influence the performance of a system.

3.2 Viola & Jones

In 2001 Paul Viola and Michael Jones proposed a framework for robust and extremely rapid object detection using optical tracking, e.g. a webcam. Nowadays, their open source framework is widely used for face detection applications.

One of the distinctions between Viola & Jones' framework and other object detectors is that it is feature based instead of pixel based. Feature based algorithms are much more efficient than pixel based algorithms. Viola and Jones use features that are reminiscent of Haar Basis functions. Because these features look somewhat similar to Haar wavelets, they are called Haar-like features.

To categorize images the feature set considers rectangular regions of the image and sums up the pixels in this region. Viola and Jones use three kinds of features. The value of a two-rectangle feature is the difference between the sum of the pixels within two rectangular regions. The regions have the same size and shape and are horizontally or vertically adjacent. A three-rectangle feature computes the sum within two outside rectangles subtracted from the sum in a center rectangle. Finally a four-rectangle feature computes the difference between diagonal pairs of rectangles.

Rectangle features can be computed very rapidly using an image representation called the integral image. The integral image at location (x,y) contains the sum of all pixels above and to the left of the point (x,y), including the pixel values in point (x,y).

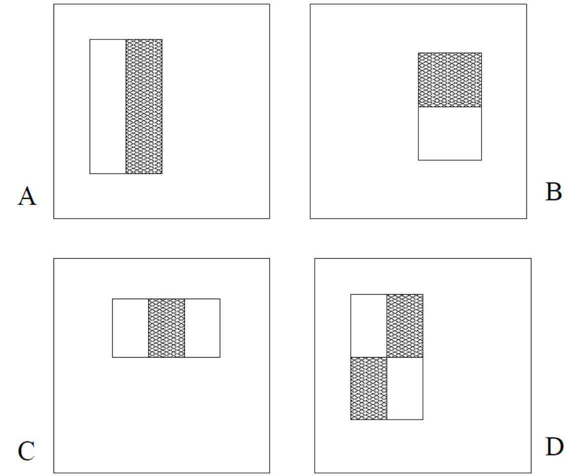


figure 3.1: Haar-like features

With the help of the integral image any rectangular area of the Haar-like features can be calculated with only four values and in constant time. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is the sum of pixels in A + B, in location 3 A + C and in location 4 A + B + C + D. The sum within D can be computed as D = ii4 + ii1 – ii2 – (3). When the sum of a calculated feature is more than a predefined threshold, the Haar-like feature is present.

Even though computing the features using the integral image is very fast, the set of possible Haar-like features in a 24 x 24 pixel sub window of the image is so large, that it would be too time consuming to calculate all of them. That is why the majority of features should be ignored and only the most important Haar-like features should be computed. Viola & Jones use a variant of the AdaBoost machine learning algorithm to select the most important features and construct a classification function.

Adaboost, which stands for Adaptive Boosting, can be seen as a heuristic that boosts classification performance. If a heuristic can improve the odds of a guess by a very small amount then it is called a weak learner. A heuristic that can improve the odds of a guess by a significant amount is called a strong learner. Boosting is a method of combining several weak learners to generate a strong learner.

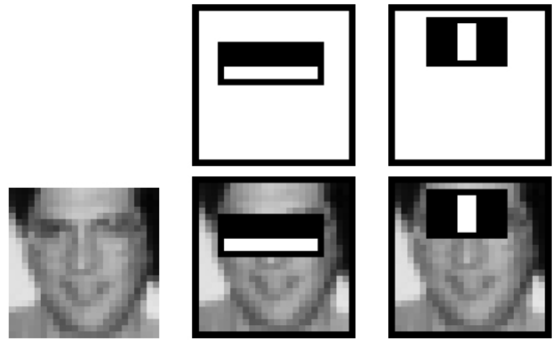


figure 3.2: Haar-like features mapped to a face

The first and second features selected by AdaBoost. The two features are shown in the top row and then overlayed on a typical training face in the bottom row. The first feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The second feature compares the intensities in the eye regions to the intensity across the bridge of the nose.

Finally an algorithm to construct a cascade of classifiers is used to increase detection performance and decrease computation time. The overall form of the detection process is that of a degenerate decision tree, a “cascade”. This cascade attempts to reject as many negatives as possible at the earliest stage possible.

The key insight Viola & Jones had is that that smaller, and therefore more efficient, boosted classifiers can be constructed which reject many of the negative sub-windows while detecting almost all positive instances. The degenerate decision tree starts with simpler classifiers that detect almost 100% of the promising features but have a false positive rate of almost 40%. In each successive step of the cascade the classifiers will become increasingly complex to cope with the increased difficulty of the classification task. As a result the false detection rate will eventually decrease to almost zero. *(Viola et al, 2001), (Torelle et al, 2009), (Silicon Intelligence, 2006), (Böhme et al, 2009)*

3.3 Eye/gaze tracking

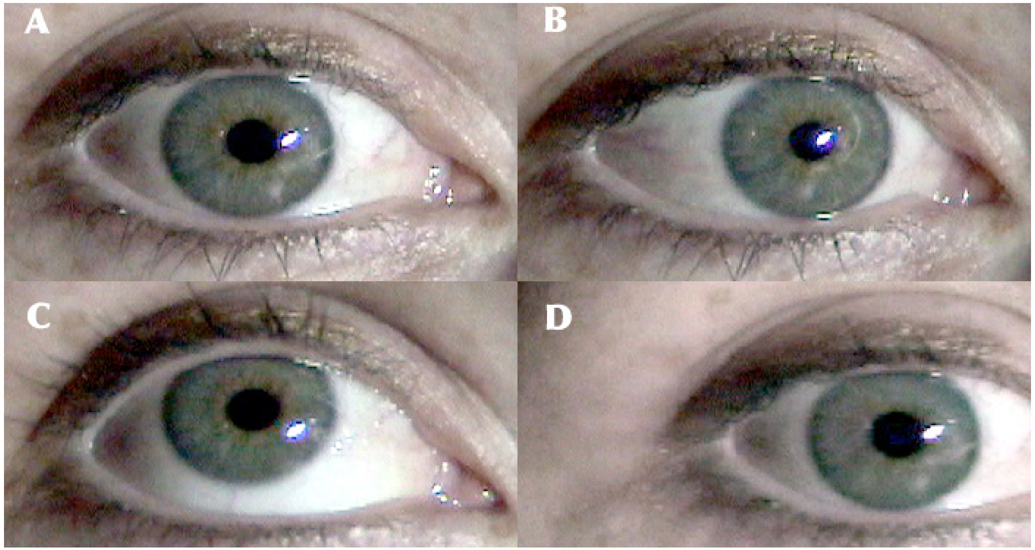


figure 3.3: The corneal reflection stays in the same position, independent of eye movements

Eye trackers measure the position of the eye or the point of gaze of a user (what the user is looking at). There are several types of tracking available. One of the techniques uses special contact lenses to measure eye movement. Another option is to place electrodes around the eye that measure the change in the electric field as a result of eye movement. The third and most widely adopted technique analyzes images of the eye by using a camera. Since the first two techniques are very encumbering, only the third will be described in detail.

In optical eye tracking the eyes are illuminated by an artificial light source that typically produces infrared light. The light reflects on the eyes and is sensed by some sort of optical sensor, ranging from a webcam to specially designed optics. The images are analyzed and eye position and rotation are determined based on changes in reflections. Light rays striking the eye produce four reflections, called Purkinje images, from the front and rear surfaces of the cornea and lens. Current trackers commonly use the center of the eye and the first Purkinje image (also known as a corneal reflection or glint) to determine the gaze.

Some methods use contrast to detect the iris to determine the center of the eye. This method performs very well in horizontal eye movements, because of good contrast between the iris and the white of the sclera. The downside is that its performance in vertical eye movement is very poor, because of eyelids obscuring the sclera and iris. A better method to determine the center of the eye is to use the pupil. In bright pupil tracking the illumination is coaxial with the optical path. The eyes reflect the light directly back towards the camera, creating a bright pupil effect, similar to the red eyes seen in photos. If the illumination is not in line with the optical path the pupil appears dark, hence the name dark pupil tracking. This happens because the reflecting light from the eye is directed away from the camera. Bright pupil tracking is more robust, because it creates greater contrast between the pupil and the iris.

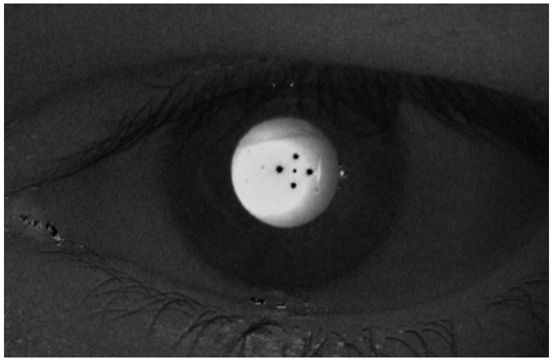


figure 3.4: Bright pupil tracking

One way to determine the position of the center of the pupil is to select all the pixels under or above a given threshold and compute the center of mass. To eliminate noise, this process is repeated in the area surrounding the previously found center, until the center stabilizes. Another way is to use an algorithm that finds the contour of the pupil. The center of the pupil is the average of all the pupil (contour) candidate points found by the algorithm. Since the position of the corneal reflection in relation to the center of the pupil remains constant during head translation, but moves with eye rotation, the point of gaze can be found.

A more accurate technique to determine the point of gaze is based on the same logic. In dual Purkinje tracking, the first as well as the fourth Purkinje image are used. Like in the method that makes use of the pupil center, these two images move similarly under translation but differentially under rotation. The change in their separation is used to determine eye rotation. The advantage of Dual Purkinje trackers is that they are very fast and accurate. A disadvantage is that they require the user's head to be stabilized, in order for the detection mechanism, which consists of a series of mirrors and servomotors, to work. *(Li D, 2006), (Li D et al, 2006), (Richardson et al, 2004), (Perez et al, 2003), (Wikipedia 2011), (Fourward Technologies 2010)*

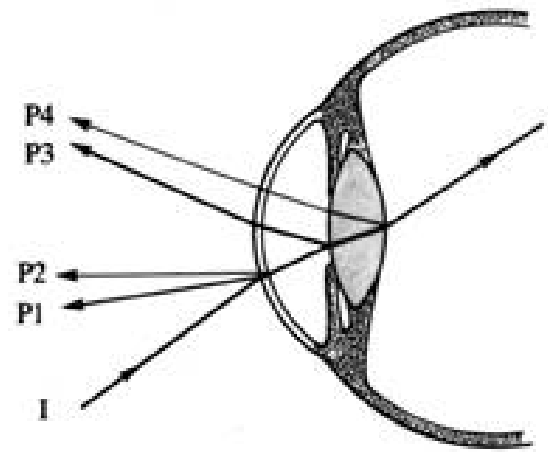


figure 3.5 The four Purkinje images

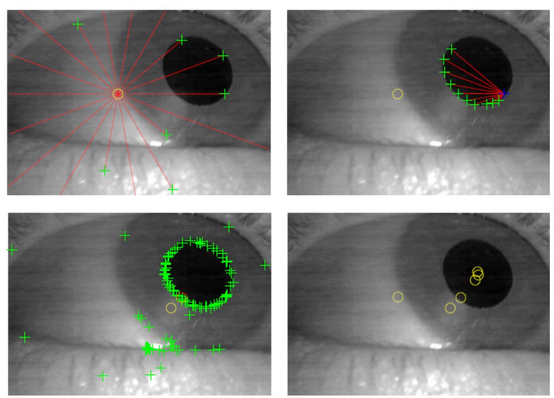


figure 3.6: Finding the contour of the pupil using an algorithm

3.4 Infrared tracking (Wii)

When the Wii was introduced in 2005 it revolutionized the way people interacted with video games. Over the years controllers had gained more functionality and better ergonomics, but this was the first time user input was analogous to movements in real life. Nintendo designed a system that can track up to four controllers using very little processing power. The wii-mote controller senses position and motion using an infrared camera and an accelerometer.

The controller features a high-end infrared camera which provides high-resolution (1024x768) images at a 100 Hz refresh rate. The camera chip has onboard multi object tracking and can track up to four IR light sources simultaneously. The Wii comes equipped with the so called sensor bar that provides two infrared dots at a known distance from each other. Five Leds are incorporated on either end of the sensor bar with one of the leds pointed slightly outwards and one slightly inwards to enhance the range of the optical sensor in the Wii-mote. From a distance the sensor sees the five individual leds as one light source.

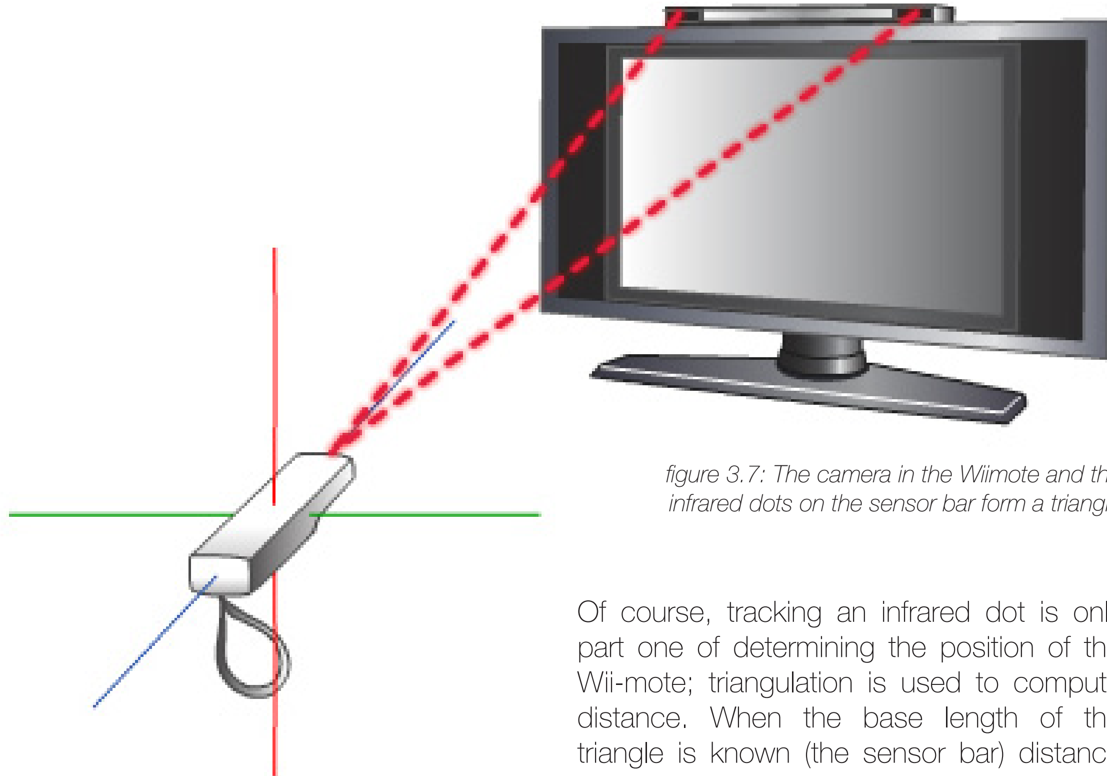


figure 3.7: The camera in the Wiimote and the infrared dots on the sensor bar form a triangle

Of course, tracking an infrared dot is only part one of determining the position of the Wii-mote; triangulation is used to compute distance. When the base length of the triangle is known (the sensor bar) distance can be computed by determining the angles from the camera to the sensor bar. The roll of the motion controller can be determined by the orientation of the dots relative to the horizontal plane.

There is a caveat in the way the wii-mote triangulates. The closer the two dots appear on the sensor, the greater is the distance from the sensor bar. However, if one would stand wayside of the sensor bar and point the remote directly at it, the dots would also appear to be very close. The distance from the sensor bar would appear to be large, while this is not necessarily the case. It is likely that Nintendo made the assumption that people would always stand somewhat in the center in front of the television, thus the effectv of looking at the sensor bar at an angle would be negligible. However, when used in head tracking applications, this could pose a problem as a result of head movements. (I.e. generating unwanted zoom effects)

(Lee J C, 2008), (LaViola Jr. J J, 2009), (Wikipedia 2011)

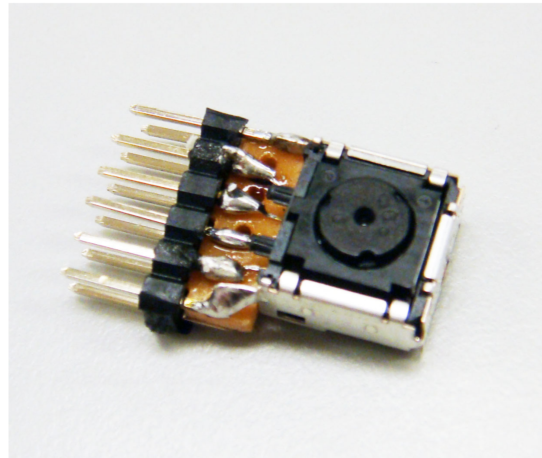


figure 3.8: The infrared camera inside the Wiimote

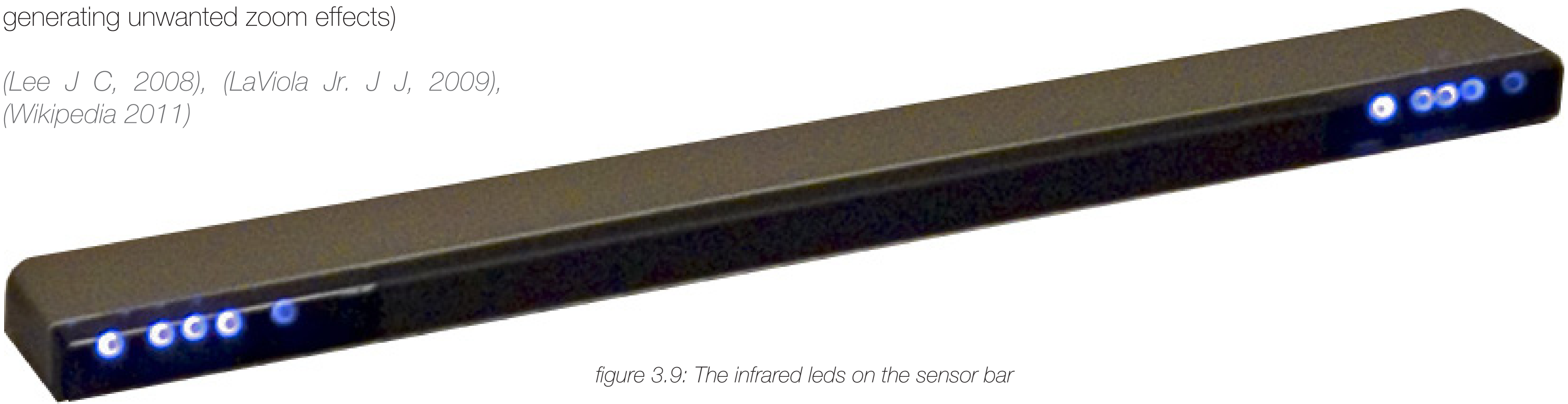


figure 3.9: The infrared leds on the sensor bar

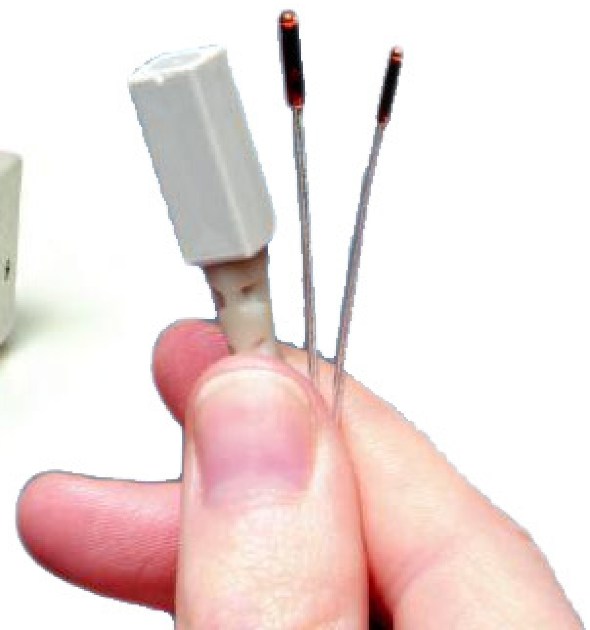
3.5 Magnetic tracking

Magnetic trackers are used to determine the rotation and position of objects in the real world.

A transmitter gives off a magnetic field, which is detected by sensors. The sensors pass this information to a filter. The filter uses the strength of the field at the sensor to compute the position and direction of each sensor, relative to the transmitter. Since the transmitter is in a fixed location, this gives the exact position and direction of each sensor.

Electromagnetic tracking has been available for about 30 years and is widely adopted in a variety of scenarios where objects need to be tracked. Magnetic trackers have the advantage that they do not require a direct line-of-sight between the transmitter and the sensor. The system has a very low latency and under ideal circumstances it is very accurate. Although magnetic tracking certainly has its advantages, it has one major drawback. When a magnetic field is transmitted, "eddy currents" are induced in conductive metals that interfere with the field transmitted by the tracker. This interference pattern affects the position and orientation outputs, resulting in distorted measurements. The distortion is even worse when the transmitter is used near ferrous metals.

figure 3.10: The Ascension trakSTAR tracker



To overcome this problem DC (direct current) tracking was invented. On paper, the well-known AC (alternating current) tracking invented by Polhemus yields better results, but in real life DC tracking can outperform AC tracking. According to Ascension, the inventor of DC tracking, the difference is the following:

AC Tracking Technology: Because of their rapidly varying nature, AC fields continuously induce eddy currents in nearby metals. Whenever conductive metal is in the tracking volume, AC trackers measurements will be distorted.

DC Tracking Technology: Pulsed DC fields reach a steady magnetic state soon after transmission. Once this condition is reached, no new eddy currents are generated. By sampling the field when eddy currents are decaying or died out, DC trackers operate with minimal or no distortion.

(Cruz-Neira et al, 1993),
(Ascension technology corp., 2009)

figure 3.11: The Polhymus Fastrak tracker

3.6 Inertial guidance system

Inertial guidance systems measure the position, velocity, acceleration and orientation of a moving object via dread reckoning. When a previously determined position of the object is known, the current position can be estimated by advancing the former position based on integrated gyroscope data and double integrated accelerometer data in time.

Rate gyroscopes measure angular velocity, and if integrated over time, provide the change in angle (or orientation) in respect to an initially known angle.

Linear accelerometers measure the vector of acceleration a and gravitational acceleration g in sensor coordinates. The sensor signals can be expressed in the global reference system if the orientation of the sensor is known. After removing the gravity component, the acceleration at can be integrated once to velocity vt and twice to position pt.

A problem that arises when integrating the sensor data is that small errors in measurement of angular velocity and acceleration errors result in progressively larger errors in velocity and even larger errors in position. To counter this "integration drift" inertial guidance systems use sensor fusion.

Sensor fusion refers to processes in which signals from two or more types of sensor are used to update or maintain the state of a system. Inertial guidance systems often use several magnetometers to reduce drift. Magnetometers measure the strength and direction of the local magnetic field, allowing the north direction to be found. While this works perfectly in some environments, the performance of a magnetometer might suffer in an environment with a large metal construction and a lot of electronics.

(Wikipedia, 2011), (Xsens Technology, 2009), (Woodman O, 2007)

figure 3.12: The xSens mtw wireless motion tracker

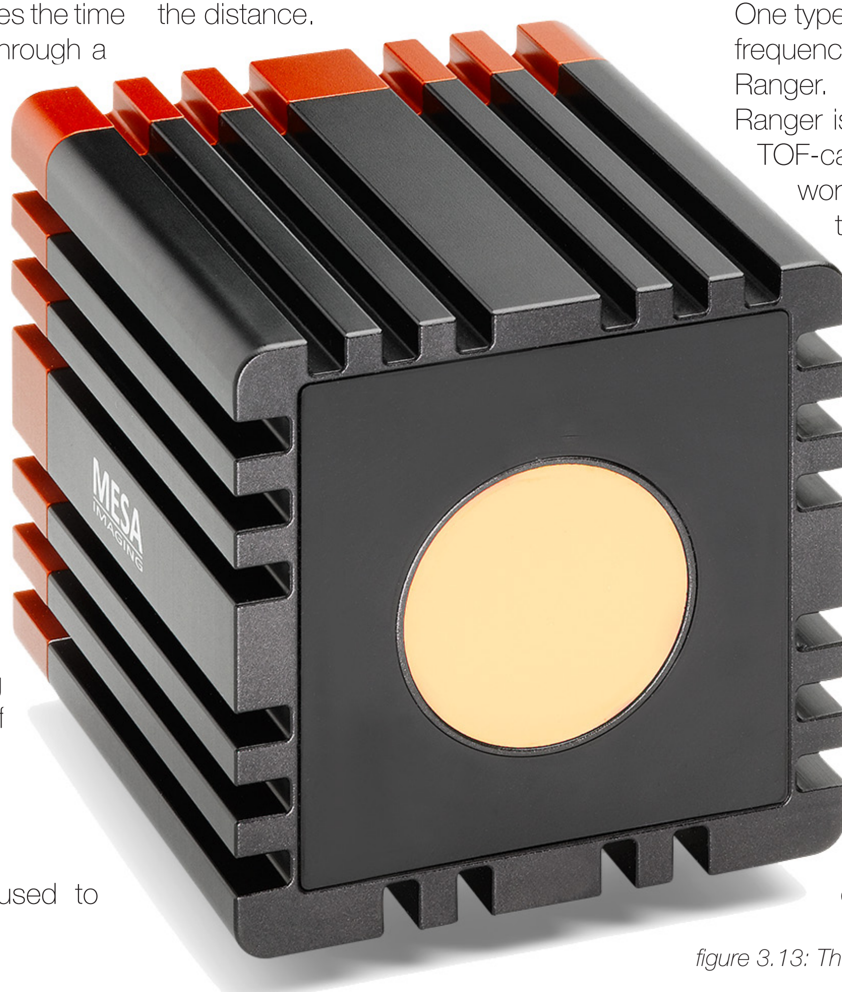
3.7 Time of flight camera

This camera system is based on the time of flight (TOF) principle, which measures the time it takes for a light wave to travel through a medium. The system emits a very short light pulse that is reflected by the objects within the illuminated scene. The camera lens gathers the reflected light and projects it on an image sensor. Each pixel in the image sensor consists of a photosensitive element (e.g. photodiode) which converts the incoming light to an electrical signal. By measuring the delay between the emission of the light pulse and the gathering of the light on the image sensor a range map can be created.

In general the distance can be calculated using the following formula, where c is the speed of light and t_D is the delay:

Distance = $1/2 \times c \times t_D$

However, various methods are used to measure the delay of the pulse and calculate the distance.



Swissranger

One type of TOF-camera is based on a radio frequency modulated light source, the Swiss Ranger. Because it seemed that the Swiss Ranger is currently one of the most popular TOF-camera's in the academic world its working principle is used to explain the technology in general. In this system a sinusoidal modulation with a frequency of 20 MHz is applied to the emitted optical signal. Each pixel in the image sensor is able to demodulate the reflected signal synchronously. The computed phase delay φ is directly proportional to the target distance and the offset B can be used to provide a 2D intensity image (which is the common input for the Viola-Jones framework).

An element that plays an important role in the distance and accuracy equations of this type of system is the so called non-ambiguity distance range. The camera can only measure without

figure 3.13: The Swissranger SR4000 time of flight camera

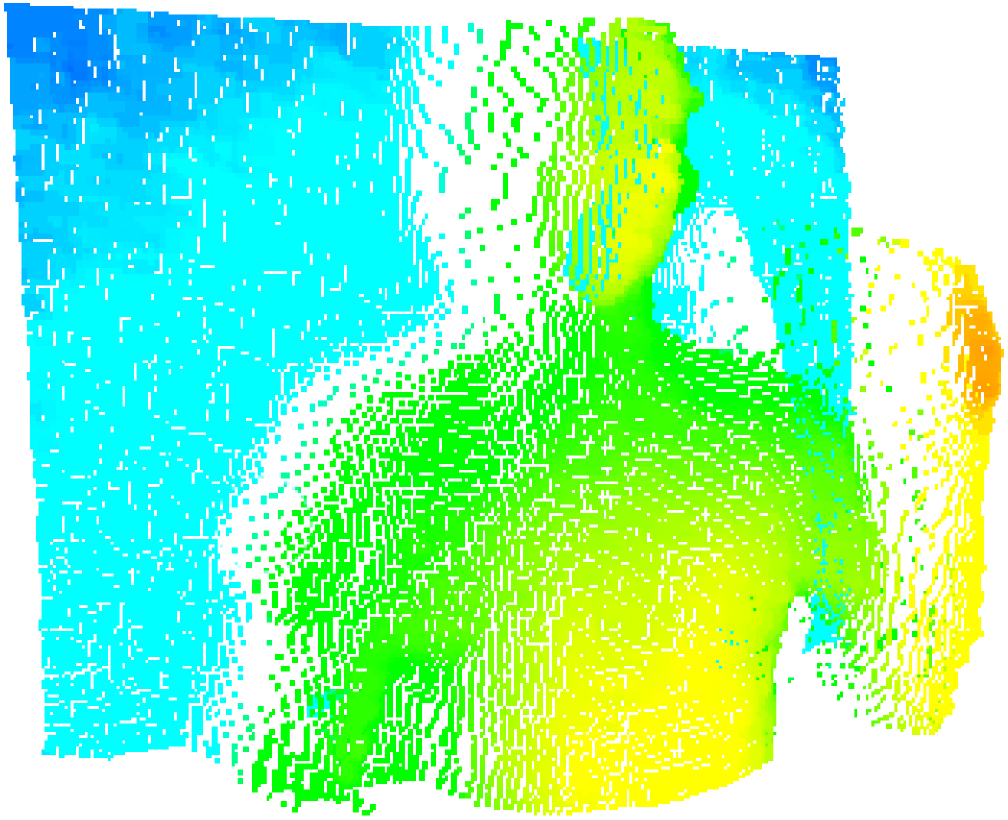
ambiguity objects that are situated in its measurement range, which is determined by the modulation frequency of the system. If an object is situated beyond the measurement range of the camera and its intensity is still high enough to be detected, the incoming light will have a delay greater than the pulse width of the system. For a camera system with a range of 7.5 meters (20Mhz) this has the result that an object at 10 meters distance will be measured at 2.5 meters and one at 16 meters distance will be measured at 1 meter.

This problem does not always occur in practice, because the range could extend further than the environment. When this is not the case the solution is to set an amplitude threshold, which filters out the signals beyond the non-ambiguity range. This works best if the objects situated in the background reflect less of the light emitted than objects in the foreground (e.g. the reflection on a very reflective surface in the background can still pass through the filter).

Another key factor that determines the performance of the camera system is the minimum incoming optical energy that needs to be detected to achieve certain distance accuracy. The theoretical distance accuracy is limited by photon shot noise. This is a type of electronic noise that occurs when the finite number of photons that carry energy is small enough to give rise to detectable statistical fluctuations in a measurement. The SwissRanger approaches the theoretical limit and thus performs very well.

In conclusion, this type of TOF-camera can simultaneously provide a distance map, an intensity image and the relative distance accuracy for each pixel within the image sensor. *(Weingarten et al, 2004), (Oggier et al, 2004), (Wikipedia, 2011)*

figure 3.14: A range/distance map



3.8 Structured light (Kinect)

The Kinect is a gaming controller that is based on a hardware design by PrimeSense and software design by Microsoft. Contrary to what was long believed, the controller is not based on the time of flight principle. Instead it uses a variant of the structured light principle, called "light Coding".

The kinect emits a near-infrared constant random speckle pattern (light code) on the scene. A laser diode continuously shines through a diffuser that produces the speckle pattern. The rays then pass through a diffractive optical element that causes them to have different focal points. This way the speckle pattern is in focus all the time, independent of the distance from the sensor.

An important difference between this system and more conventional structured light systems is that a constant random pattern is used instead of a periodic pattern. This prevents the so called "wrapping problem", where movements larger than the period of the projected pattern cannot be distinguished. In that case it would be ambiguous in what period a movement took place. This problem is similar to the non-ambiguity range in time of flight cameras.

The reflecting light is sensed by an infrared CMOS sensor that looks at the scene from a different angle. When the speckle is reflected by an object, the pattern differs from that of a reference image stored in the Kinect. Besides a difference in scale, the speckle pattern does not vary in the axial direction that passes through the object (Z-axis). Objects cause the pattern to shift in the transversal plane, i.e. the detection plane.

A "prediction-based region-growing" correlation algorithm is used to compute depth. The method first tries to find new region anchor points with a correlation value higher than a certain threshold, i.e. they correlate with the reference image. The depth can be calculated using triangulation, because the distance between the emitter and the camera is known, as well as their angles. When these anchor points are found region growing is applied and the depth of the neighboring points is predicted, using the assumption that depth does not change much within a certain region. If there is a small difference in depth, the point is joined to the region of the anchor point. If the difference in depth is too large, the point probably belongs to a different region.

(Nongenre, 2011),
(Bits & Chips, 2011),
(Shotton et al, 2011),
(Prime Sense, Ltd., 2007),
(Wikipedia, 2011),
(RPC Photonics, 2011)

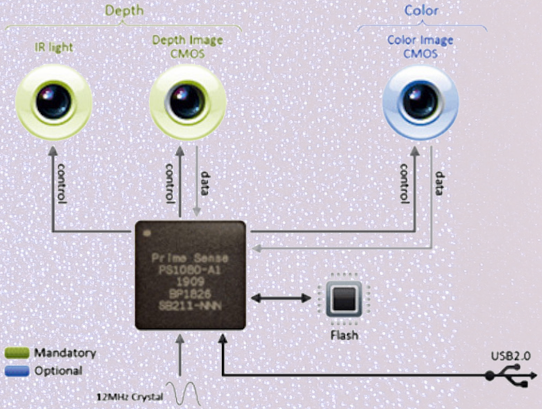


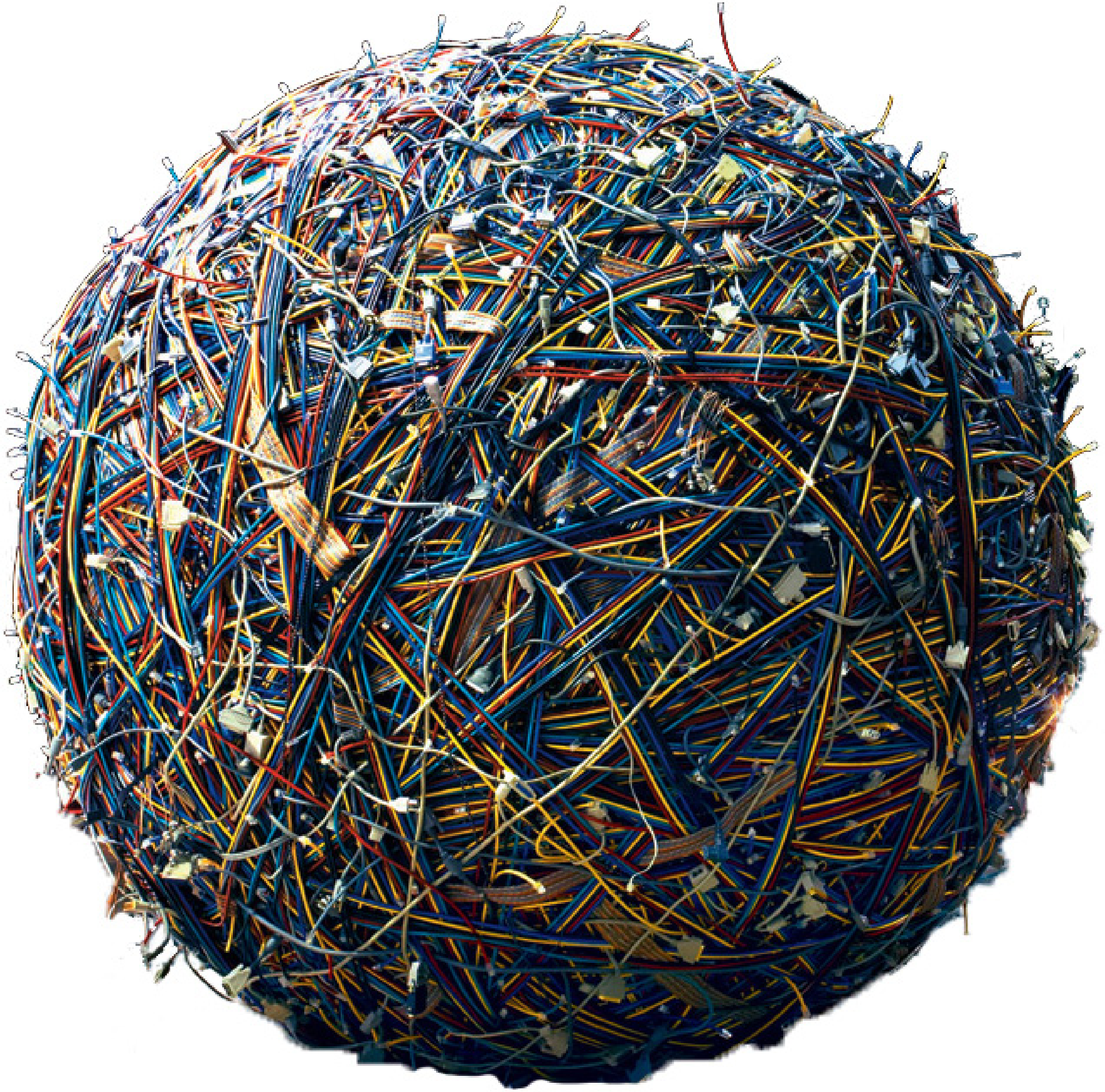
figure 3.15: How the Kinect works

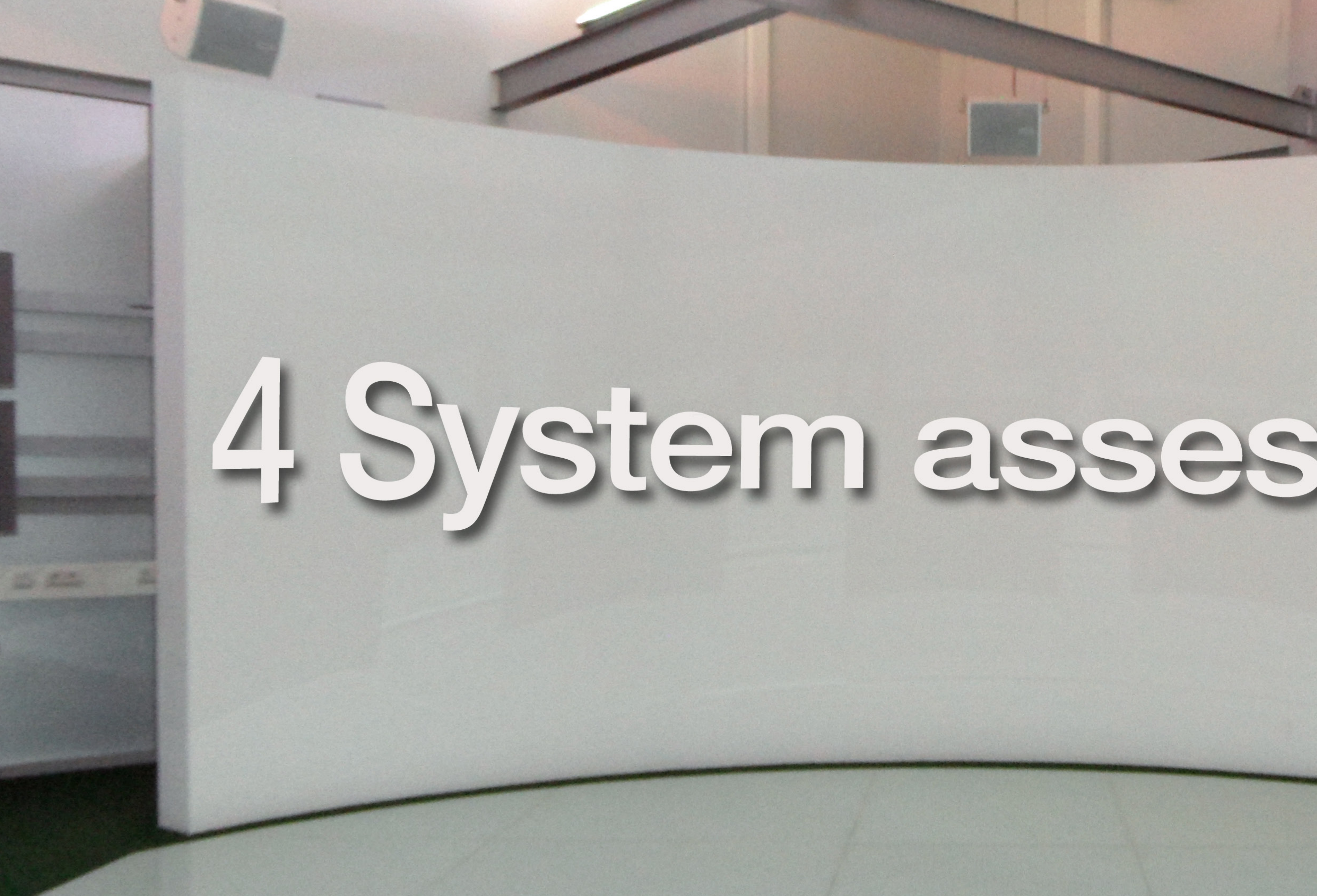
figure 3.16: Inside the Kinect

3.9 Conclusion

In theory, every technology detailed in this chapter is capable of delivering headtracking that is accurate and has low latency. In practice however, not every tracking technology is equally suitable to use in the VR-laboratory. Because there are many variables that play a role in the performance of a tracking method, there is no ideal technology that stands above the crowd. Especially the environment can have a great effect on the performance. The metal construction, abundance of electronics and difficult lighting conditions (beamer lights, screens) make the VR-lab a harsh environment to operate in for most of the technologies. Other factors influencing the choice for a certain technology are the price of necessary equipment and comfort of use (encumbering versus unencumbering to use).

To make a grounded decision on what technology the final result of this thesis should be based on, more variables should be taken into account. In the next chapter the variables that play a role within the VR-lab are delved into, and made clear.





4 System asses

4.1 Introduction

During the exploration of scientific papers on perspective corrected displays and headtracking it became apparent that many variables influence the performance of a system. Some tracking technologies perform magnificent under certain conditions, while performing well below par in other conditions. This chapter tries to clarify most of the parameters that might play a role in the (tracking) decision making process. Finally a tool to assess what the solution route should be is discussed.

4.2 VR applications

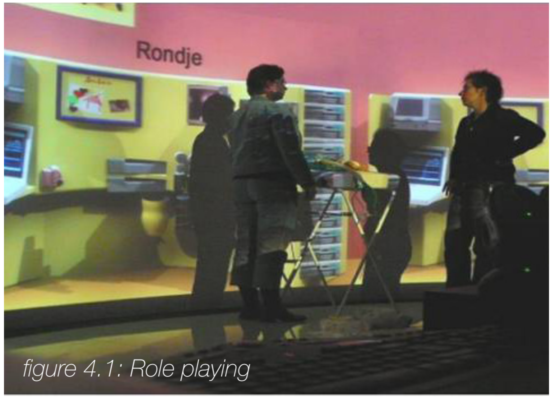


figure 4.1: Role playing

figure 4.2: Edugame

figure 4.3: Virtual "Maasvlakte"

4.2.1 ROLE PLAYING

In order to explore design concepts with users or to assess a design at an early stage in the design process, role playing can be used. Research shows that prior acting experience is not necessary for a role playing session to be successful. However, it can sometimes prove difficult to imagine a concept design within a certain scenario. Virtual reality can be used to create an immersive environment that supports the user in envisioning a concept within a scenario. In the VR-lab the theater wall has been used multiple times for this purpose. *(Svanæs et al, 2004)*

4.2.2 EDUGAME

In some, if not all, design projects it is important to involve stakeholders actively. It can be difficult to involve people with different interests, expertise and or professional language. Edugames can provide a framework for participatory design that helps in such a way that everyone involved can make design moves and be part of the exploring and negotiating process in order to create common images of possible futures. Edugames are played with multiple users, and require that facilities can be used simultaneously. *(Brandt E, 2006)*

4.2.3 EXPLORE VIRTUAL ENVIRONMENTS

The VR-lab offers the possibility to explore existing or future environments virtually. In Rotterdam, the visitor information center for the second Maasvlakte used a large virtual environment to present what this area will look like when it is finished in 2033. While this serves a more communicative and entertaining purpose, exploring a virtual environment can also be used in a more professional sense. To assess design decisions, a virtual representation of the environment can be built in which the user can navigate. *(DPI Animation House, 2009)*

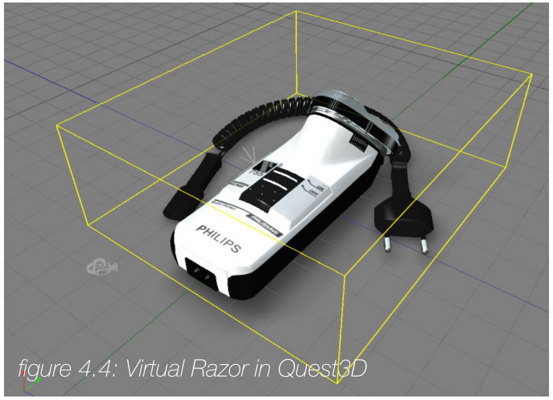


figure 4.4: Virtual Razor in Quest3D


figure 4.5: Ship simulator


figure 4.6: Incident management training

4.2.4 EVALUATE VIRTUAL OBJECT

Besides exploring virtual environments, it is also possible to evaluate the objects one can find within such an environment. How does a new design fit into its surroundings? Or is it practical to use? VR can offer more than just a three dimensional view of an object. It can enable the user to interact with a virtual object, by means of haptic devices or gaming devices like the Microsoft Kinect and the Wiimote. When interacting with an object it is important that the user's proprioceptive information is analogous to the motion on screen.

4.2.5 SIMULATION

Simulation is one of the applications where virtual reality can save a lot of time and expenses. Simulator systems are used for testing, training, research and demonstrational purposes. While the VR-lab does not incorporate purpose built simulators, like flight simulators, it offers a broad range of possibilities with its multitude of screens and input devices. One of the applications used in the VR-lab simulates driving a car on the highway. *(VR-lab University of Twente, 2011)*

4.2.6 INCIDENT MANAGEMENT TRAINING

It is very easy to set up a certain scenario in a VR environment and it is no wonder that this feature makes VR a popular tool for incident management training. Once a month members from several emergency services meet in the VR-lab to train for possible incident scenarios. They have to assess risks and dangers, decide which measures to take and what procedures to apply, and report to the other rescue crew members.

4.3 Screen types

figure 4.7: Gaming tables

figure 4.8: Driving sim on theather wall

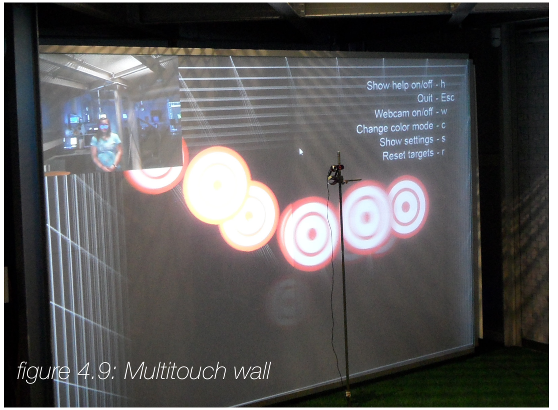


figure 4.9: Multitouch wall

4.3.1 SERIOUS GAMING TABLES

The serious gaming tables are mainly used to support group activities. There are 9 tablet computers available, 3 projection screens (side by side) and one projection table. Cameras are used to support touch recognition of all projection screens. An example of a group activity the gaming table is used for is disaster training where members from different emergency services play out a disaster scenario.

4.3.2 THEATHER WALL

The theatre wall consists of two beamers and a large curved projection screen, measuring 8 meters in width and 3 meters in height. It is very suitable for simulations and role playing because it delivers an immersive experience. As a perspective corrected display it benefits from the curved projection screen, which makes it possible to look around, without going past the boundaries of the screen. A downside of the theatre wall is that it currently only operates in a Windows XP environment (video drivers). This can cause compatibility issues with newer equipment (e.g. a webcam).

4.3.3 TOUCH WALL

The touch wall is a custom made screen that is based on the technology of the surface table. It consists of a rear projection screen, two beamers and a series of camera's that detect touch. The screen measures approximately 3.5 by 2 meters end extends all the way to the floor. The screen is immersive because the user can stand at very close range, without occluding the imagery.

4.3.4 MICROSOFT SURFACE

Surface is a 30-inch display in a table-like form factor that small groups can use at the same time. It was developed by Microsoft to create technology that would bridge the physical and virtual worlds. The table that went to market in 2007 supports direct interaction, multitouch and object recognition. Because of its horizontal form factor and intuitive interaction it is mainly used as an input device for multiple users. *(Microsoft, 2007)*

figure 4.10: Microsoft Surface 1.0

4.3.5 ELUMENS VISION STATION

The Elumens Vision Station is a VR-dome that offers a 180 degrees field of view. The dome takes over your entire visual field, including your peripheral vision. This makes it a very immersive experience that is ideal for simulating environments. It consists of hemispherical projection screen and a modified beamer with a special lens that enables 180 degrees projection. A benefit of the Elumens Vision Station is that you are immersed in the experience without losing contact with your surroundings. *(How Stuff Works, 2011)*

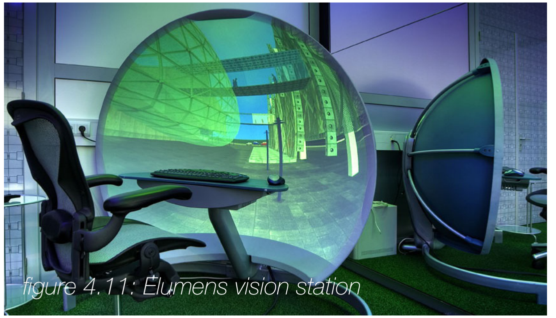


figure 4.11: Elumens vision station

4.3.6 AUTOSTEREOSCOPIC LCD MONITOR

When it comes to 3D technology, the VR-lab is mainly interested in glasses free 3D. Current consumer screens rely mostly on 3D enabled by active shutter glasses or polarized glasses, but the VR-lab houses an auto stereoscopic screen from Philips. Multiple users can simultaneously watch stereoscopic images, without having to wear anything. The technology used in this particular screen is still in its infancy, and the 3D effect is not very natural.

4.3.7 LEGACY SCREENS

Besides exotic screens, there is an abundance of legacy screens available in the laboratory, ranging from tablets or desktop screens to large lcd screens. While they don't have particularly interesting features, these screens are very suitable for desktop virtual reality as demonstrated by Johnny Lee (Wii desktop VR) and Seeing Machines (FaceApi).

4.4 Users

Obviously, a perspective corrected view can only be offered to one user at a time, which is inherent in the principle. However, multiple screen setups like the game tables can offer side-by-side perspective corrected views. This can be important to prevent undermining the group experience some virtual reality applications rely on. It remains untested at this moment if side-by-side views have a positive or negative effect on the use of an application.

Many different users visit the VR-lab, of which some are highly experienced with VR while others might not even be familiar with a computer. In finding a good technology to enable head tracking, one has to take into account how well a user can operate the technology and interact with an application. For example, novice users can find it difficult to calibrate a certain technology or feel inclined to move in an unnatural way to achieve a shift in perspective on the screen.

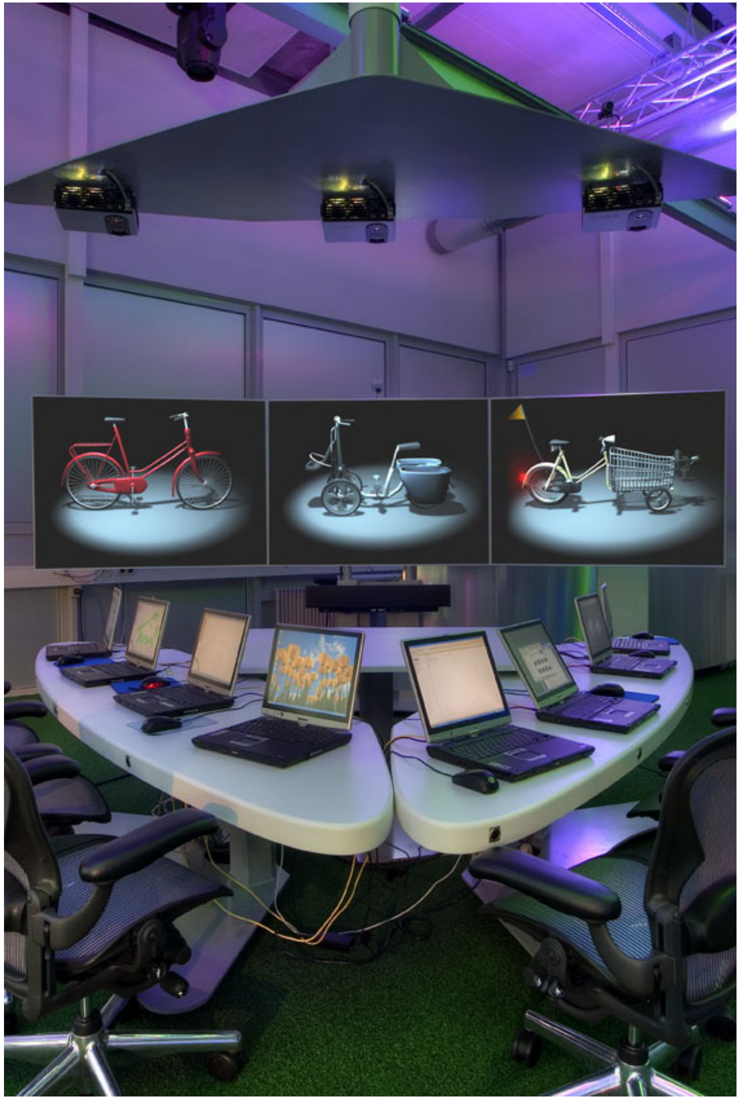


figure 4.12: Multiple users can use the game tables

4.5 VR-lab: Environment

The environment within the lab is optimized for VR use and uses a steel structure to suspend many of the electronics (e.g. beamers, lights). The light within the VR-lab can be controlled, but it is usually quite dark inside, to optimize screen brightness. A large amount of people can be accommodated, but it is also common that there is only one researcher/user at work at a time. There is a significant amount of electronics present of which some devices are always active, or in standby. It is likely that all these devices generate quite some electronic noise. Also, bright spots and heat sources are present, because of beamers and spotlights.

All this can give a tracking system a difficult time. A dark environment has a negative effect on optical tracking systems, while heat sources can negate the benefits of infrared tracking. A large metal frame and electronic noise can impair magnetic tracking and inertial navigation. When there are multiple groups present in the lab, even a time-of-flight camera or a Kinect might find it difficult to find the actual user.

figure 4.13: The VR-lab (T-Exchange) of the University of Twente

4.6 Tracking methods

The tracking technologies that are taken into account in this Bachelor thesis are discussed in depth in chapter three. Because they play an important role in the assessment tool and to recapitulate their main benefits and drawbacks are mentioned here.

4.6.1 VIOLA & JONES FACE TRACKING

Viola and Jones offer a very robust framework that can detect faces with low latency while using relatively small computational power. The user does not need to wear anything, making it a very unencumbering system. The only hardware and software needed for the system to work is a webcam, a computer and the Open CV library (Open Source Computer Vision library).

However, the framework is susceptible to lighting noise, which is present in the VR-lab. Multiple users can occlude each other and cause the system to track the wrong user. Also, because there are a lot of objects present in the lab they might invoke false positives.

4.6.2 EYE TRACKING/GAZE TRACKING

Depending on the technology used for eye or gaze tracking, it can be just as unencumbering as optical tracking based on Viola and Jones framework. Besides a computer and a webcam, eye/gaze tracking can benefit greatly from an artificial (infrared) light source. The most interesting improvement upon standard face tracking is the ability to track the gaze of the user. Translations of the head as well as eye movement will change the perspective on the screen, giving a more natural experience.

The downside is that this method has a very limited range, thus the camera has to be very close to the user. Also, the technology is very dependent on lighting conditions; this can lead to an unstable system as a result of the many bright spots in the lab.

4.6.3 INFRARED TRACKING (WII REMOTE)

The Wiimote might be last year’s tech, but it is still a very clever system. It is easily available for a low price and there is a large user base that generates open source software and code. Since the tracking is done within the remote, it requires only little computational power. The main advantage of using a Wiimote for tracking is its robustness. It only tracks infrared light dots and the sensitivity can be adjusted so that it only sees the brightest ones.

A negative aspect of using a Wiimote for tracking is that it is less accurate than the other methods. Also the user is required to wear either the Wiimote itself, or leds attached to some kind of framework (e.g. glasses, headphones). The Wiimote was never intended to be used on a computer and that has its disadvantages. It will only work on a limited variety of operating systems and does not support all Bluetooth stacks.

4.6.4 WIRED MAGNETIC TRACKING

This type of tracking can be used in any lighting condition, which is a very positive point when we take the VR-lab in mind. Also, it does not suffer from occlusion caused by multiple users in front of the same screen. Theoretically this tracking method could provide a perspective corrected view on all screens in the VR-lab at once! It has a large range and the user should be able to walk freely inside the lab, while still being tracked in six degrees of freedom.

Unfortunately there are two things that this type of technology cannot overcome. Contrary to more traditional buildings, the VR-lab consists of a large steel framework that can distort measurements. Obviously, the VR-lab is loaded with electronics, which will distort measurements as well. Also, it is not very unencumbering to use.

4.6.5 INERTIAL GUIDANCE SYSTEM

The inertial guidance system is somewhat similar to magnetic tracking in that it has a large range and can track six degrees of freedom. Also it does not suffer from occlusion or lighting problems. Besides the sensor, it only needs a computer to work. Another benefit is that one of the leading companies that develop these sensors is a spinoff of the University of Twente, and sensors are already available.

The user still has to wear a sensor, although it is quite small. Also, the sensor needs to be calibrated before use. The technology suffers from cumulative tracking errors because it uses dread reckoning to determine the position of the sensor. Normally so called “sensor fusion” can overcome this, by combining the sensor data with a magnetometer. But, as stated before, this type of sensor functions poorly in an environment that has a metal construction and a lot of electronics.

4.6.6 TIME OF FLIGHT CAMERA

The time of flight camera is still primarily used within a scientific context, but it has promising features that make it very suitable for other application areas as well. Just like optical trackers that use a webcam, it is an unencumbering technology that does not require the user to wear anything. A full distance map of the environment is created within the sensor, thus little computational power is needed for tracking. The main advantages of a time of flight camera are that it is relatively insensitive to its surroundings and that it achieves very high accuracy.

For once, the downside of this tracking method does not have to do with technical hurdles. Time of flight cameras are still expensive (depending on the budget) and the amount of different commercially available products is limited. Perhaps more important is the fact that it does not offer a complete tracking solution. The depth map can make tracking much easier and more robust, but it does still need a tracking algorithm.

4.6.7 MICROSOFT KINECT

The technology in Microsoft’s Kinect has been reverse engineered and seemed to resemble the time of flight principle at first. However, since the inner workings are different, their positive and negative points vary. The sensor resembles a webcam on steroids and is very unencumbering. Microsoft was able to mass market its Kinect and it is offered at an unparalleled price point for this type of sensor. It is not susceptible to magnetic or electronic noise and generally handles different lighting conditions well.

The Kinect’s success is largely based on a smart algorithm that requires moderatly intensive computations at this moment. Although the pattern used for sensing is in focus all the time, the system has a limited field of view and the sensor needs to be tilted (motorized) to enhance it. One other downside of this system is that it can only be calibrated in the factory. If something changes in the diffractive optical element, resulting in a slightly altered pattern, measurements will be distorted.

4.7 Assessment tool

In the previous paragraphs all the variables that are taken into account are discussed. While all the information is useful on its own, one quickly loses overview. That is why all variables should be combined in an assessment tool which can help to decide what route to take in terms of perspective corrected displays.

Some of the variables can influence each other. E.g. not all screens are suitable for certain VR applications and they will only support a perspective corrected view to a limited amount of users simultaneously. The tool is built upon Microsoft Excel and cross references all data.

A user selects the VR application, the screen type, the number of simultaneous users and the type of user. Finally a tracking method is chosen and the user selects what elements are present in the environment (e.g. metallic structure). Based on the input given by the user, the Excel document will show all problems that could arise and need attention. It is fairly easy to add more parameters to the document that will increase detail or expand the research area.

figure 4.14: The front-end of the assessment tool. A user follows the steps and ends with selecting elements which are present in the environment. The issues that could arise are shown in red.

1. Application	2. Screen type	3. Number of simultaneous users	4. Type of user	5. Tracking method	6. Situa
Role_playing	Reality_theather	2	Beginner	Eye_tracking	Dark_e
	Sensor placement	Performance decrease when views are connected	Inefficiency	Occlusion	Decrea
		Computational intensive		Sensor needs to be close	Decrea
					Decrea

In case of the VR-lab at the University of Twente a few things stand out in the assessment. The metal construction and large amount of electronics will render magnetic tracking and inertial navigation systems practically useless. Furthermore the difficult lighting conditions as a result of bright spots and a dim ambience (to increase screen visibility) decrease optical tracking performance. An overall good performer is the time of flight camera, since it is relatively insensitive to most of the conditions of the VR-lab environment. In that respect the Wiimote and Kinect are solid performers too, but they have some shortcomings. The Wiimote requires the user to wear infrared leds and the Kinect might be difficult to place in front of the larger screens, because of its dimensions.

The assessment tool is build in Excel and is intended to give insight in what such a tool could look like and how it should function. A final tool should be very dynamic, but that would require some object oriented programming, which needs a larger timeframe than available in this thesis. The Excel sheet primarily shows the core functionality.

ation			
nvironment	Light_environment	Bright_spots	Metal_construction
x	x	x	x
ses face tracking performance	No issues	Decreases face tracking performance	
ses eye tracking performance		Decreases eye tracking performance	
ses gaze tracking performance		Decreases infrared tracking performance	

*=ALS(F$3=”X”;ALS(VERSCHUIVING(‘
xREFTracking-Situation’!A1;VERT.
ZOEKEN(E2;’5 Tracking
method’!$A:$B;2;ONWAAR);VERT.
ZOEKEN(F$2;’6 Situation’!$A:$B
;2;ONWAAR))=WAAR;ALS(VERT.
ZOEKEN(F1&F$2&RIJ()-5;’Issues
(2)’!$A:$E;5;ONWAAR)=””;””;VERT.
ZOEKEN(F1&F$2&RIJ()-5;’Issues
(2)’!$A:$E;5;ONWAAR));””);””)*

4.8 Conclusion

Application	Reality_theather	Multitouch_wall	Game_table	TV_screen	VR_Dome	Microsoft_surface_table
Role_playing	WAAR	WAAR	ONWAAR	ONWAAR	ONWAAR	ONWAAR
Edugame	WAAR	WAAR	WAAR	WAAR	WAAR	WAAR
Explore_virtual_environments	WAAR	WAAR	WAAR	WAAR	WAAR	WAAR
Evaluate_virtual_object	WAAR	WAAR	WAAR	WAAR	WAAR	WAAR
Simulation	WAAR	WAAR	WAAR	WAAR	WAAR	ONWAAR
Incident_management	WAAR	WAAR	WAAR	ONWAAR	ONWAAR	ONWAAR

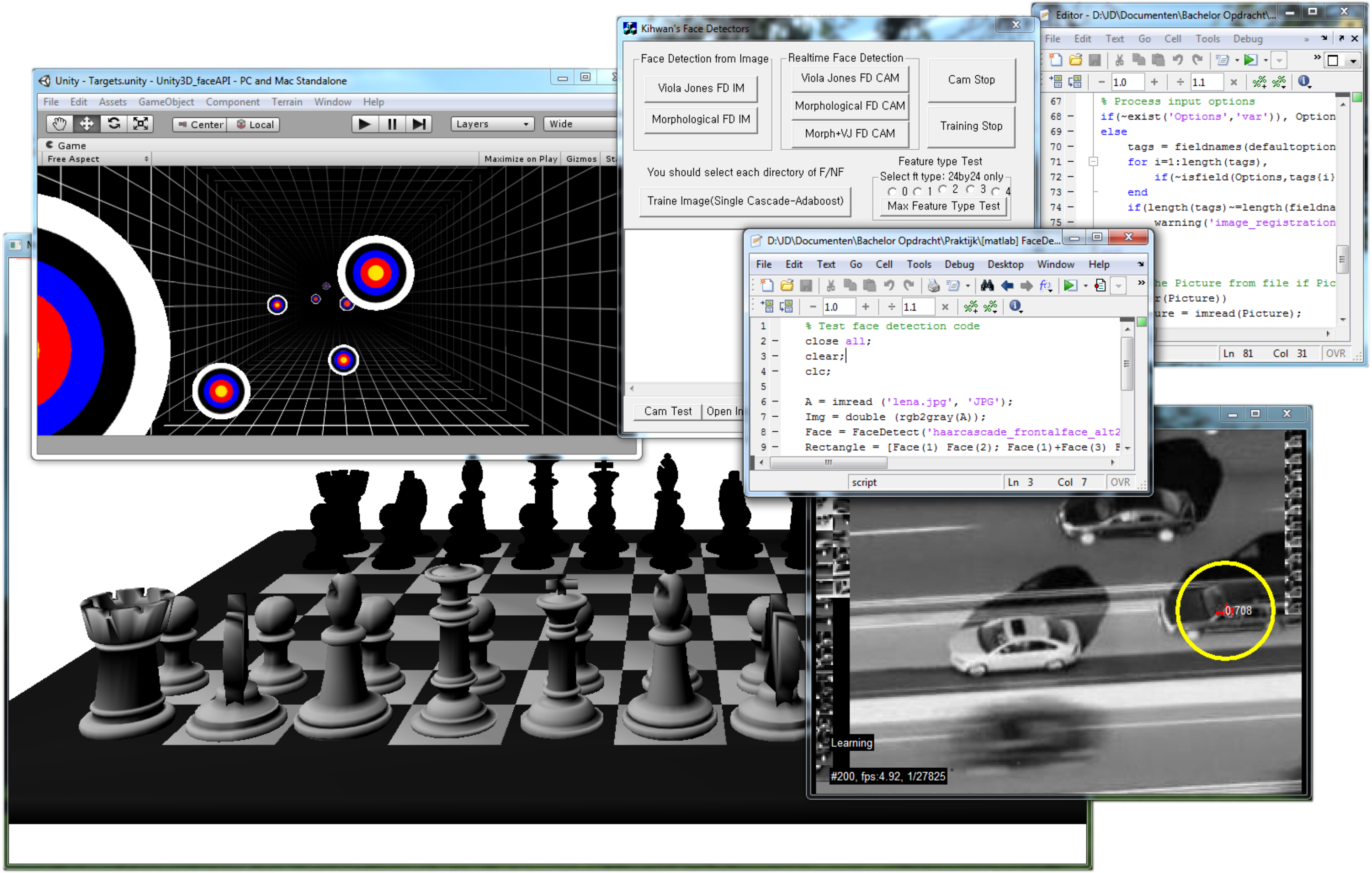
figure 4.15: Cross referencing is used to filter out incompatible elements

The excel document is a helpful tool to maintain overview of all parameters and assess what solution route is best. Still, the more detail is available in the document, the more grounded a choice of tracking technology will be. Currently, some of the issues described in the tool, might be too broadly formulated to be helpful to a novice user. Someone who has experience in this field would be more aware of nuances that aren't immediately apparent in the tool. Specific demands of the VR-lab, like excluding the use of glasses, did not find their way into the tool.

Based on the findings of the author and the assessment tool, the Time-of-flight camera emerges as the (theoretically) best solution. However, not all solution routes will fit into the scope of this thesis. There is no TOF-camera available at the University at this moment and acquiring one would be too expensive to justify for testing purposes only. That is why a pragmatic solution has to be found for a proof of principle of perspective corrected displays.

The next chapter will focus on some preliminary testing, to find out what would be a realistic solution route that fits within the scope of the thesis and delivers a perspective corrected display in practice.

figure 4.16: One of the formulas used in the Excel document


Unity - Targets.unity - Unity3D_faceAPI - PC and Mac Standalone
File Edit Assets GameObject Component Terrain Window Help
Center Local
Layers
Wide
Game
Free Aspect
Maximize on Play Gizmos
Kihwan's Face Detectors
Face Detection from Image
Viola Jones FD IM
Morphological FD IM
Realtime Face Detection
Viola Jones FD CAM
Morphological FD CAM
Morph+VJ FD CAM
Cam Stop
Training Stop
You should select each directory of F/NF
Traine Image(Single Cascade-Adaboost)
Feature type Test
Select ft type: 24by24 only
Max Feature Type Test
Cam Test
Editor - D:\JD\Documenten\Bachelor Opdracht\...
File Edit Text Go Cell Tools Debug
% Process input options
if(~exist('Options','var')), Option
else
tags = fieldnames(defaultoption
for i=1:length(tags),
if(~isfield(Options,tags{i}
end
if(length(tags)~=length(fieldna
warning('image_registration
Ln 81 Col 31 OVR
D:\JD\Documenten\Bachelor Opdracht\Praktijk\[matlab] FaceDe...
File Edit Text Go Cell Tools Debug Desktop Window Help
% Test face detection code
close all;
clear;
clc;
A = imread ('lena.jpg', 'JPG');
Img = double (rgb2gray(A));
Face = FaceDetect('haarcascade_frontalface_alt2
Rectangle = [Face(1) Face(2); Face(1)+Face(3)
script Ln 3 Col 7 OVR
0.708
Learning
#200, fps:4.92, 1/27825

5 Preliminary testing

5.1 Introduction

A lot of knowledge has been gained about head tracking and variables that play a role in the VR-lab. To find viable candidates that could be implemented in the VR-lab and that fit the scope of this thesis a short testing phase was carried out. The goal of this phase was to assess which tracking technology was readily available and what application/ technology would perform best. The key was to test-drive applications without spending too much time on it.

Out of the technologies discussed in chapter three and four, only applications based on the Wiimote or a webcam were considered. The Kinect was not available for purchase at that time and the other technologies were already deemed unsuitable for one or more reasons.

5.2 Wii

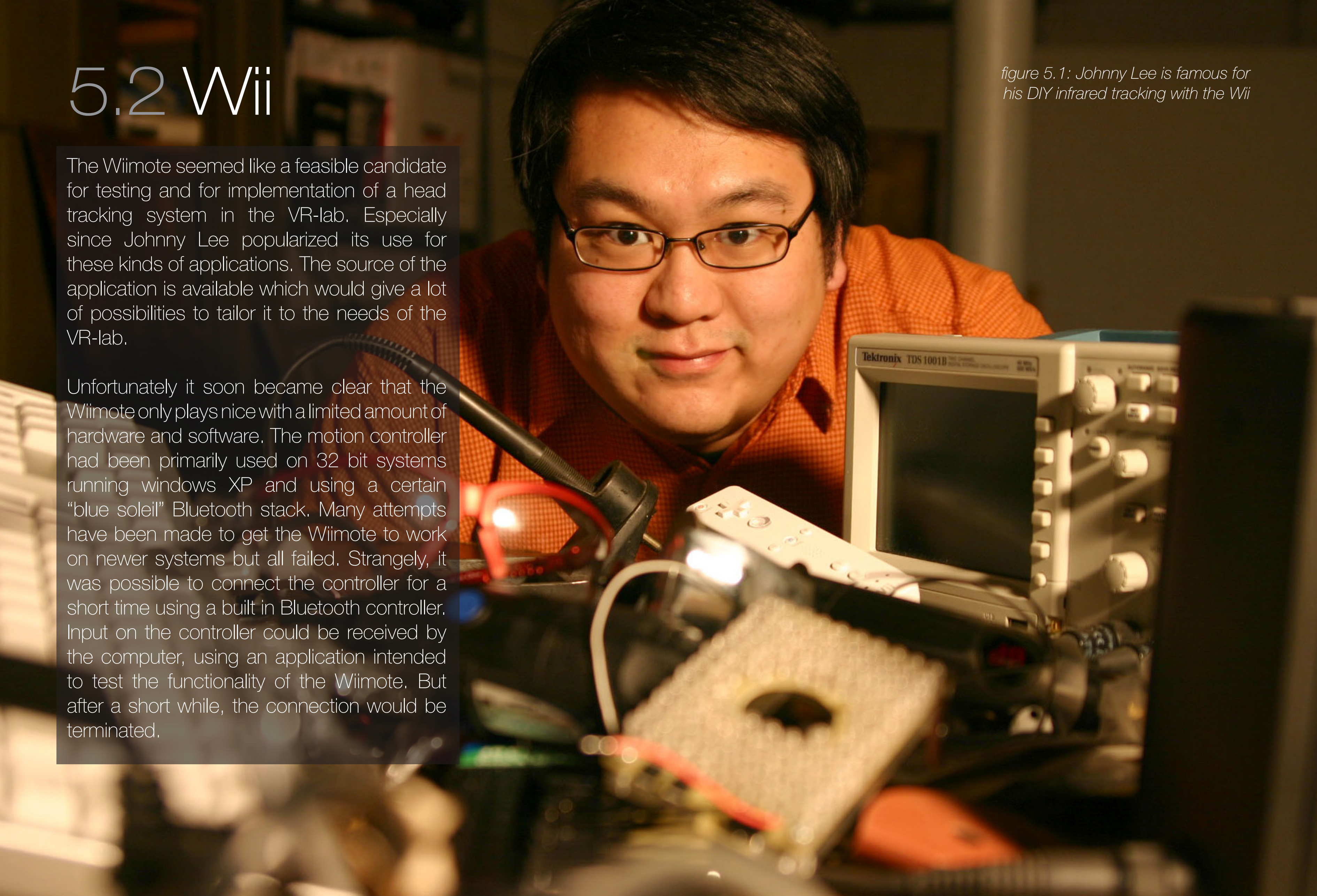


figure 5.1: Johnny Lee is famous for his DIY infrared tracking with the Wii

The Wiimote seemed like a feasible candidate for testing and for implementation of a head tracking system in the VR-lab. Especially since Johnny Lee popularized its use for these kinds of applications. The source of the application is available which would give a lot of possibilities to tailor it to the needs of the VR-lab.

Unfortunately it soon became clear that the Wiimote only plays nice with a limited amount of hardware and software. The motion controller had been primarily used on 32 bit systems running windows XP and using a certain “blue soleil” Bluetooth stack. Many attempts have been made to get the Wiimote to work on newer systems but all failed. Strangely, it was possible to connect the controller for a short time using a built in Bluetooth controller. Input on the controller could be received by the computer, using an application intended to test the functionality of the Wiimote. But after a short while, the connection would be terminated.

5.3 Webcam

There are many applications and/or scripts available that rely on the Viola & Jones face tracking algorithm. Many have been tried, but most where unsuitable to use in the testing phase. Several scripts suffered in performance due to bad implementation of the algorithm. Others showed good performance, but were severely limited in the possibilities because data could not be passed to other programs. And there are also applications that seem to work great, but ask a premium price for using it. Coincidently, a group of students from electrical engineering was working on a face tracking application for a study course. This paragraph is based on that application since they provided it for testing purposes and it seemed suitable.

The webcam has been tested on some of the most defining screens in the VR-lab. It is difficult to change parameters that are linked to screens like the theater wall or the game table. If a method works on these screens, it will most likely work on legacy screens (e.g. laptops, desktops, TV's).

5.3.1 THE APPLICATION

The application is based on an open source project called "magic vision portal". It uses a webcam and the openCV library, which includes Viola & Jones' algorithm, to create a perspective corrected display. No data can be passed on to other programs in this specific application, but it does a good job demoing the technology. Besides tracking capabilities, it also offers a VR-box environment like the one seen in the well-known Johnny Lee's experiments. *(Magic Vision Portal, 2011)*

The webcam should be at eyelevel for a good result. Otherwise, it can prove too difficult to find the eyes and get a reliable result without too many false positives. The application works best if a user first keeps his or her head still in front of the webcam until there is a fix. Then, more distance can be created between the webcam and a user, up till about three meters. At one meter distance from the webcam the application follows the user's motion rather fluidly. Too some extend it will also react if the user turns his or her head.

5.3.2 THEATER WALL

The performance of the application is deteriorated because of the surroundings of the webcam. The theater wall uses two large beamers to create very large, unified imagery. Because of the bright lights, the user appears quite dark and details are less visible, making it harder for the algorithm to track faces. Also, two metal beams that support the control room generated false positives in the application. By shielding the webcam from the beamer light performance increased, but the metal beams were still an issue. A different issue is that the computer that powers it resides behind the screen. The distance from the computer to the sensor is larger than the USB 2.0 specifications support (5 meters due to maximum round trip delay).

Seeing the perspective shift as you move is an impressive view in front of the theater wall. It is important that the user stands somewhere in the middle, because the curved projection screen makes the imagery appear out of proportion if a user stands on one of the sides. The application would still work if the user was standing some meters away from the webcam.

figure 5.2: Testing on the theather wall

5.3.3 MULTITOUCH WALL

Testing on the multitouch wall was primarily difficult because of the darkness in front of the screen. The webcam had difficulties distinguishing the user from the surrounding, since the image looked almost opaque. Performance increased dramatically, by using a spotlight that enlightens the user and the surrounding. The spot was placed in such a way, that no light would directly hit the webcam or blind the user.

The screen of the multitouch wall extends all the way down to the floor which gives a very natural effect. It is more forgiving in use then the theater wall, but it has to give in some immersion. Sensor placement is relatively easy, since the computer on which the application runs is nearby. Unfortunately, it occludes a small part of the screen.

5.3.4 GAME TABLE

The application works quite well on the game table. Although it uses beamers just like the theater wall, they are not nearly as bright and don’t seem to cause too much decrease in performance. Something that became apparent during testing is that the application often sees horizontal and vertical metal construction beams as faces. This is most likely due to the inner workings of the algorithm. The construction beams resemble the so called “haar-like” features that are used to distinguish elements of the face. A uniform background around the face could improve performance.

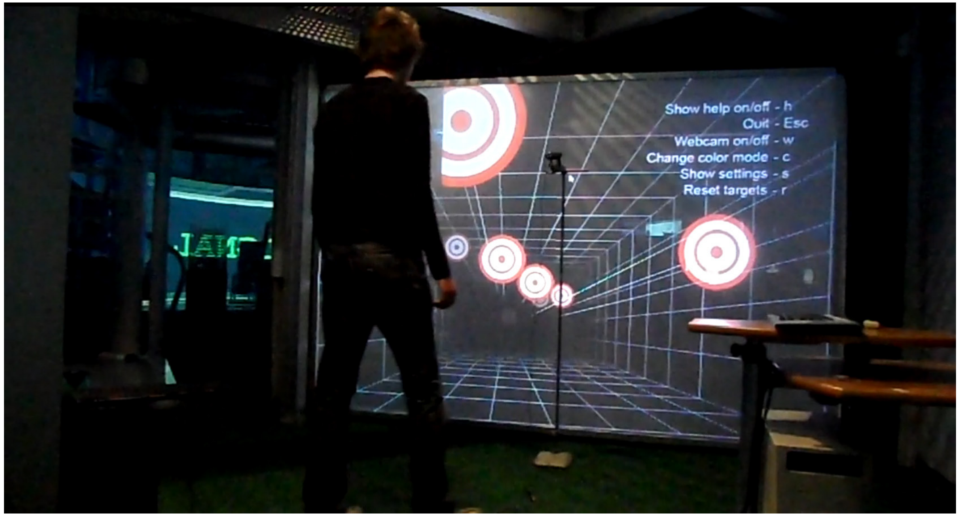


figure 5.3: Testing on the multitouch wall

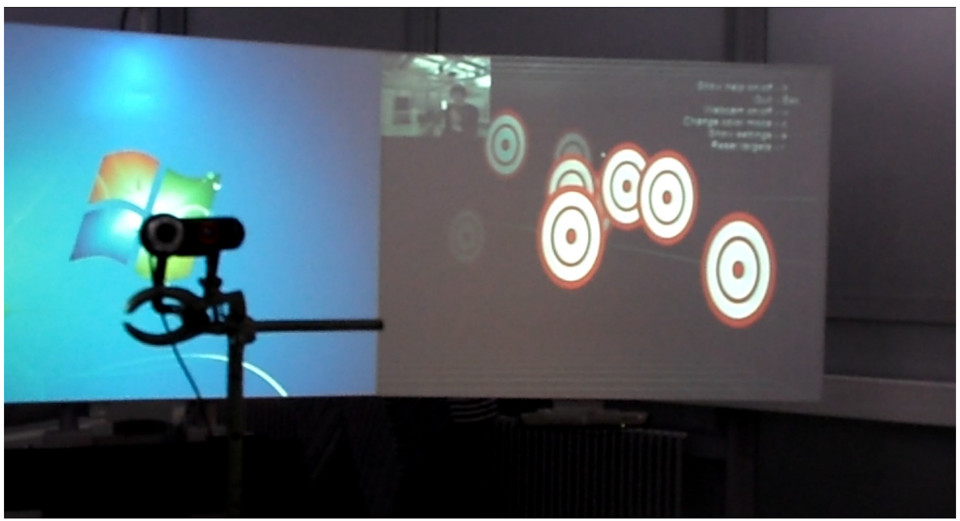
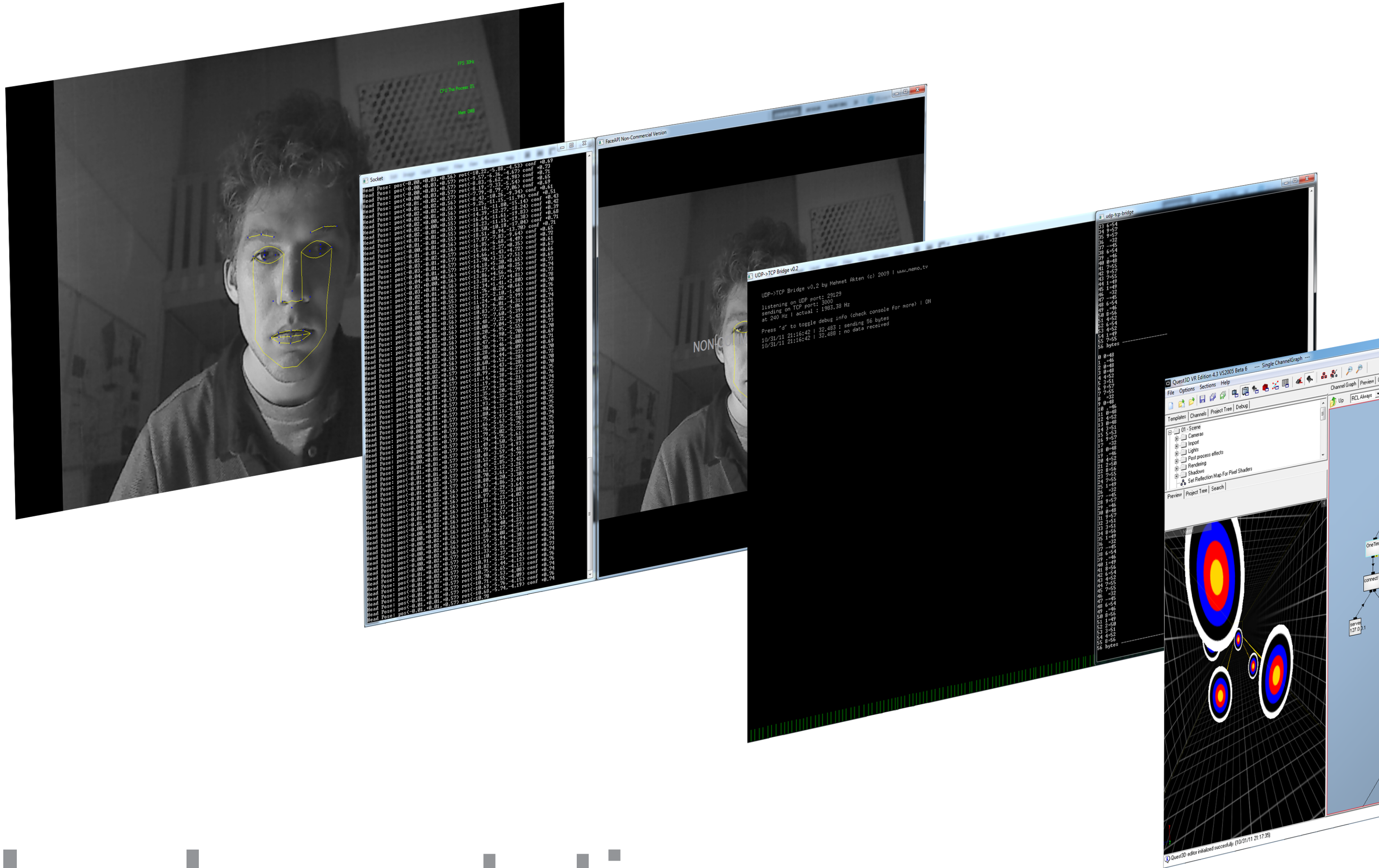


figure 5.4: Testing on the game table

5.4 Conclusion

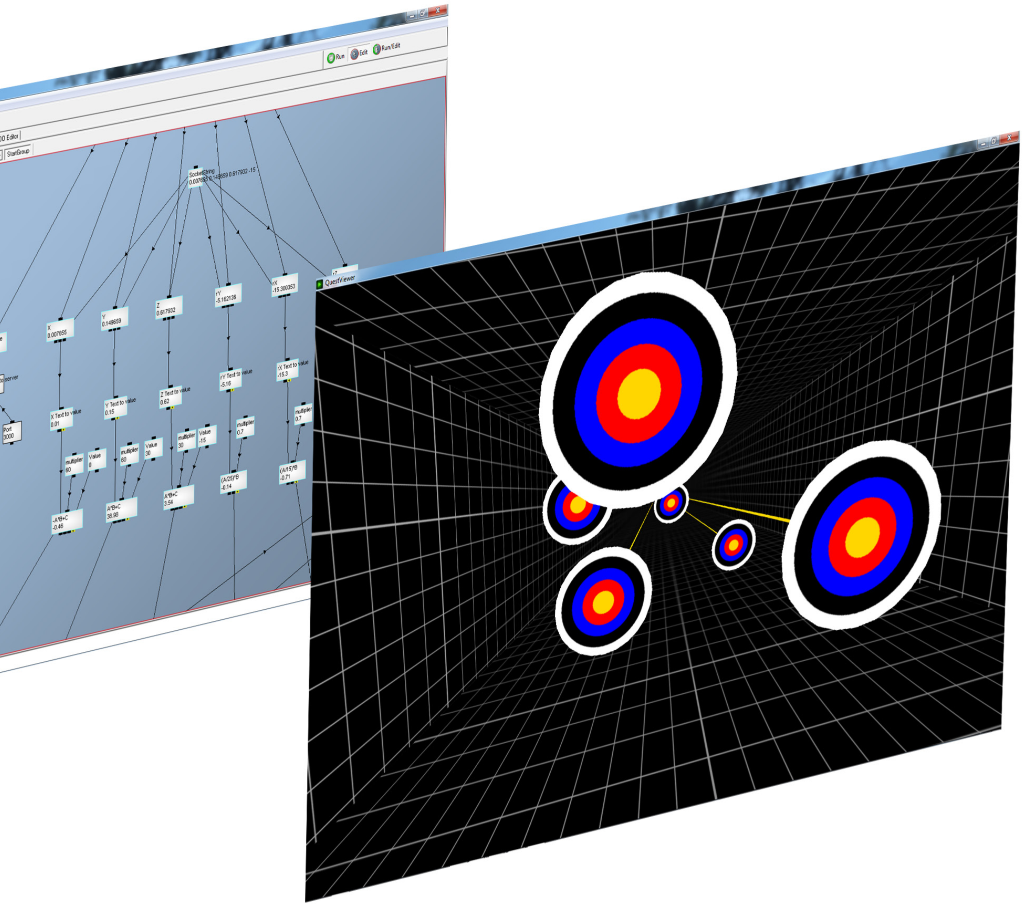
Out of the many tracking methods described in chapter three, only one seems suitable for implementation. Most methods go beyond the scope of this thesis, be it for economic reasons and/or because they need a larger timeframe. Tracking based on a webcam or a Wiimote is readily available on the internet and both seem like a pragmatic way of delivering a proof of principle of a perspective corrected display in the VR-lab. However there are a lot of stability-issues with DIY applications and the Wiimote would not even work on the computers in the lab. The test with the webcam delivered promising results, but different software is a necessity for the implementation phase.

The data acquired by the head tracking application needs to be passed on to the program Quest 3D. This program is responsible for most of the virtual environments in the virtual reality lab. The data imported in Quest 3D will be used to create a corresponding change in perspective in the VR environment.



6 Implementation

6.1 Introduction



After preliminary testing it became clear that tracking based on a webcam is the way to go, simply because there is no viable alternative available that also fits within the scope of this thesis. The implementation phase consists of further research into webcam based face tracking and a solution route. More research was needed to find an application that would enable robust headtracking and a way of passing on the acquired data to Quest3D. Many applications have been tried, most of them based on Matlab. Since none of the Matlab-based trackers yielded satisfactory results in terms of usability for the implementation phase, they will not be described below. What will be described is a solution route based on an application called FaceApi. As opposed to most of the other trackers tested, this is not open source software, but it is available in a non-commercial version.

6.2 Step one: FaceApi

The most important step towards a perspective corrected display in the VR-lab is a robust tracker that can pass data to a VR-environment, which in this case is Quest3D. A tracker that can deliver this is made by Seeing Machines Ltd. They have been working on camera-based face tracking for more than eight years, and it shows. The first versions used stereo-camera rigs and were originally designed for enhancing robot-human interaction. It then became part of Seeing Machines' "faceLAB", a research tool for studying human behavior under real-world conditions. They first used it for the detection of drowsy and inattentive drivers, which is why the tracking has to be very robust.

With "FaceApi" that technology is now readily available for download, albeit with closed source. An application programming interface (API) can be seen as a building block that is designed to communicate with other software. Quest 3D does not support headtracking natively, but by communicating with the FaceApi building block it can still be outfitted with it.

The API will work with any webcam and tracks the user's head position in Cartesian coordinates relative to the camera (X,Y,Z) and the user's head pose in Euler angles (rads). Especially the ability to track the head pose of the user is interesting, because it enables more natural motion in front of a smaller screen. I.e. the user does not have to translate his or her head necessarily.

There are no whitepapers available on the tracking technology, thus it is hard to describe how it works. However, since Seeing Machines Ltd. Claims to have a database of thousands of faces and resources of the Api indicate the use of cascades, it is very likely that it is built upon the Viola & Jones framework.

A non-commercial version is available for use in research that is not compensated financially. This version is free of charge and offers solid 6-degree-of-freedom-tracking (DOF), but has limited features. As opposed to the developer version, it does not offer lens calibration, network output and it uses a simpler version of the tracking algorithm. More technical and licensing information on FaceApi can be found in Annex C.

(Seeing Machines, 2011), (Wikipedia, 2011)

figure 6.1: FaceApi tracking


FPS: 30Hz
CPU This Process: 0%
Mem: 0MB
seeingmachines
www.seeingmachines.com

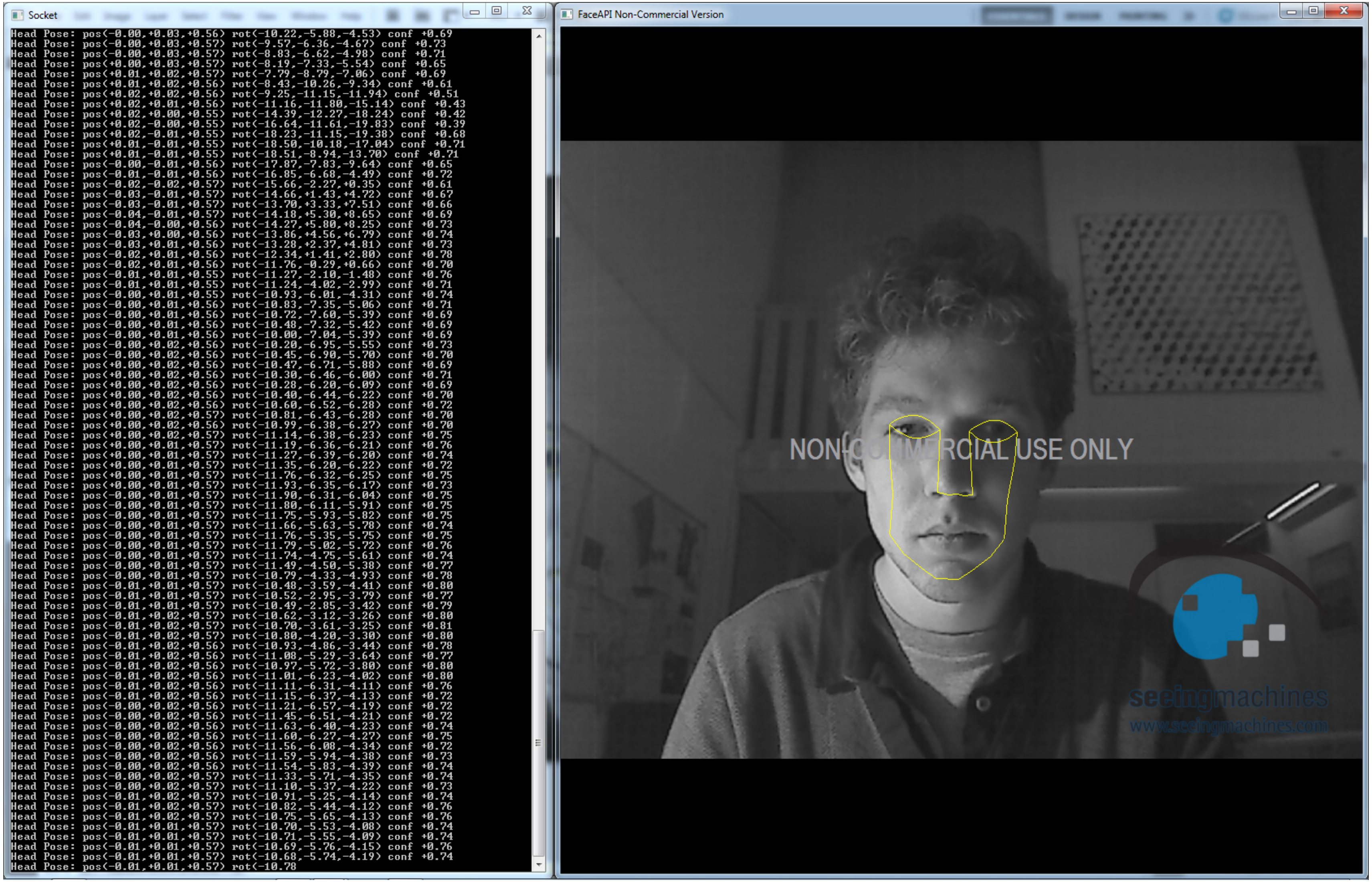
6.3 Step two: FaceApi streamer

With only the non-commercial version of the FaceApi at hand it does not seem possible to call upon the API directly from within Quest 3D. To feed the tracking data into Quest 3D it is necessary to set up some kind of data stream that Quest 3D can import. In the commercial version of FaceApi the acquired tracking data can be send over a network using an UDP stream. That option is not available in the non-commercial version, but luckily there are developers that have taken an interest in using (free) FaceApi for their applications. On an open source project called the “6dofstreamer” the FaceApi streamer can be downloaded in which the FaceApi resources (tracking capabilities) are incorporated. It uses the first webcam it finds, applies facetracking and streams all the data on the network based on UDP.

UDP is the abbreviation of User Datagram Protocol, which is one of the basic protocols on the internet. The main feature of UDP in comparison to TCP is that it is faster, because it has less overhead. It is mainly used in applications where fast transmission of data and fast response is important, like media-streams and online gaming. Unfortunately, the tradeoff can be big; an UDP stream is quite unreliable and there is no ordering applied to the packets that are send. UDP is designed to be lightweight and assumes reliability is not an issue for the application it is used for. Thus, it does not check if a packet actually arrives and packets might not arrive in the order they were send.

(6DOFstreamer, 2011),
(Wikipedia, 2011)

figure 6.2: FaceApi streamer


Socket
FaceAPI Non-Commercial Version
seeingmachines
www.seeingmachines.com

6.4 Step three: UDP-TCP bridge

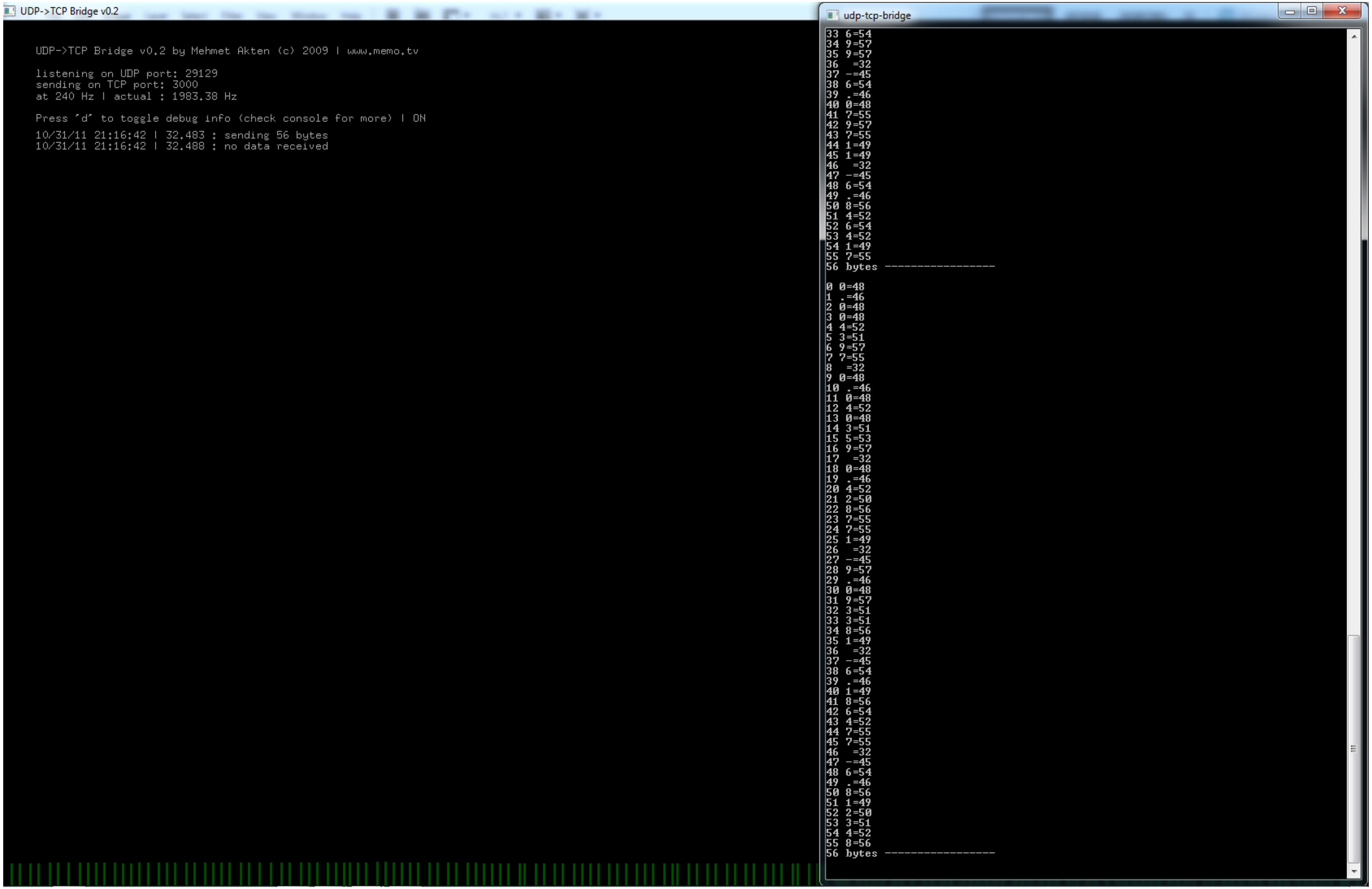
There is no native support for UDP streams (and perhaps for good reasons) in Quest 3D. This functionality can be added with so called channels that are somewhat similar to API's. There is only one channel available from a user that can listen to an UDP-stream, but that is not free of charge.

(Godbersen H, 2011)

A different approach is needed to import the data from the stream into Quest 3D. The application does support streams based on TCP and thus the UDP-stream should be converted to a TCP-stream. The open source application UDP-TCP bridge does just that. The application is set up to listen to a specific port on which it will receive the tracking data send with the UDP protocol. The application then resends the data to a preconfigured port according to the transmission control protocol (TCP).

(UDP-TCP-bridge, 2011)

figure 6.3: UDP to TCP streamer


UDP->TCP Bridge v0.2
UDP->TCP Bridge v0.2 by Mehmet Akten (c) 2009 | www.memo.tv
listening on UDP port: 29129
sending on TCP port: 3000
at 240 Hz | actual : 1983.38 Hz
Press 'd' to toggle debug info (check console for more) | ON
10/31/11 21:16:42 | 32.483 : sending 56 bytes
10/31/11 21:16:42 | 32.488 : no data received
udp-tcp-bridge
33 6=54
34 9=57
35 9=57
36 =32
37 -=45
38 6=54
39 .=46
40 0=48
41 7=55
42 9=57
43 7=55
44 1=49
45 1=49
46 =32
47 -=45
48 6=54
49 .=46
50 8=56
51 4=52
52 6=54
53 4=52
54 1=49
55 7=55
56 bytes ----------------
0 0=48
1 .=46
2 0=48
3 0=48
4 4=52
5 3=51
6 9=57
7 7=55
8 =32
9 0=48
10 .=46
11 0=48
12 4=52
13 0=48
14 3=51
15 5=53
16 9=57
17 =32
18 0=48
19 .=46
20 4=52
21 2=50
22 8=56
23 7=55
24 7=55
25 1=49
26 =32
27 -=45
28 9=57
29 .=46
30 0=48
31 9=57
32 3=51
33 3=51
34 8=56
35 1=49
36 =32
37 -=45
38 6=54
39 .=46
40 1=49
41 8=56
42 6=54
43 4=52
44 7=55
45 7=55
46 =32
47 -=45
48 6=54
49 .=46
50 8=56
51 1=49
52 2=50
53 3=51
54 4=52
55 8=56
56 bytes ----------------

6.5 Step four: Quest3D

Quest 3D is a real-time 3D development tool that is currently one of the backbones of the VR-lab of the University of Twente. In the tool, an interactive virtual environment can be build, cad-models can be imported and the finished production can be rendered in real-time.

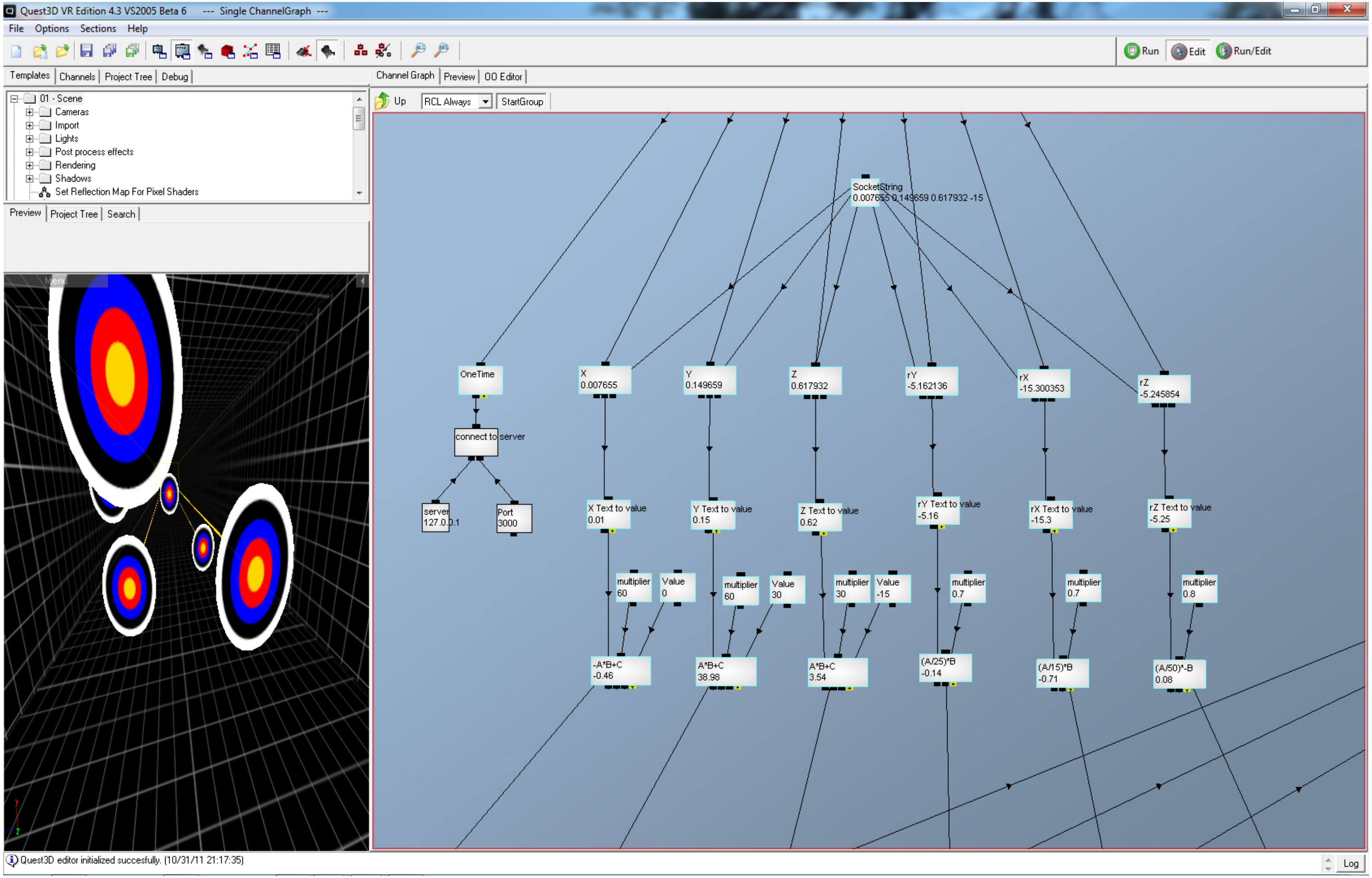
Quest 3D is developed by the Dutch company Act-3D in Leiden, an early competitor in the field of real-time 3D graphics. A real-time 3D development tool was made for in-house use. This eventually led to the development of the product now called Quest 3D in the year 2000. What is special about the tool is that it does not require the user to have any experience with programming. According to Act-3D, Quest 3D is based on a unique channel technology. A real-time 3D production is divided into components, which are displayed on the screen graphically. To create interactive functionality, the components can be connected using drag-and-drop, without any programming or scripting. Each component is reusable for other projects.

While this functionality is very useful for novice users, it does have some limitations in terms of controllability and efficiency for more advanced users (with programming experience). Also, the workspace of Quest 3D quickly becomes cluttered and/or indistinct. For example, it is difficult to link a component within a collapsed channel to a channel higher in the hierarchy.

(Act-3D, 2011)

On the next pages follows a short description of the 'quest' starting with importing data into Quest 3D and ending with camera movement induced by head movement of the user.

figure 6.4: Developing in Quest 3D


Quest3D VR Edition 4.3 VS2005 Beta 6 --- Single ChannelGraph ---
File Options Sections Help
Run Edit Run/Edit
Templates Channels Project Tree Debug
Channel Graph Preview OO Editor
01 - Scene
Cameras
Import
Lights
Post process effects
Rendering
Shadows
Set Reflection Map For Pixel Shaders
Preview Project Tree Search
Up RCL Always StartGroup
SocketString
0.007655 0.149659 0.617932 -15
OneTime
connect to server
server
127.0.0.1
Port
3000
X
0.007655
Y
0.149659
Z
0.617932
rY
-5.162136
rX
-15.300353
rZ
-5.245854
X Text to value
0.01
Y Text to value
0.15
Z Text to value
0.62
rY Text to value
-5.16
rX Text to value
-15.3
rZ Text to value
-5.25
multiplier
60
Value
0
multiplier
60
Value
30
multiplier
30
Value
-15
multiplier
0.7
multiplier
0.7
multiplier
0.8
-A*B+C
-0.46
A*B+C
38.98
A*B+C
3.54
(A/25)*B
-0.14
(A/15)*B
-0.71
(A/50)*-B
0.08
Quest3D editor initialized succesfully. (10/31/11 21:17:35)
Log

6.5.1 ACQUIRING DATA

The first step to take is to get all the data into Quest 3D, which can be more difficult than it seems as a novice user. This part of 'the quest' is based on work by Roy Damgrave, who is one of the few people with practical Quest 3D experience at the University of Twente at this moment.

The tracking system will be a separate group so it can easily be used as a building block in other quest projects. To make this group part of the whole Quest 3D project a "channelcaller" is used. This will call the children within the group.

A connection is set up with a server with the IP Address and the port number on which the data stream is transmitted. The stream transmits the data as a single string in which the position and rotation vectors are separated by a space. A SocketString channel listens to the stream and updates itself when a new string is received from the connection. The received string will be copied so it can be used within Quest 3D. The string is then filtered to separate the individual vectors and the now separate strings are converted to values.

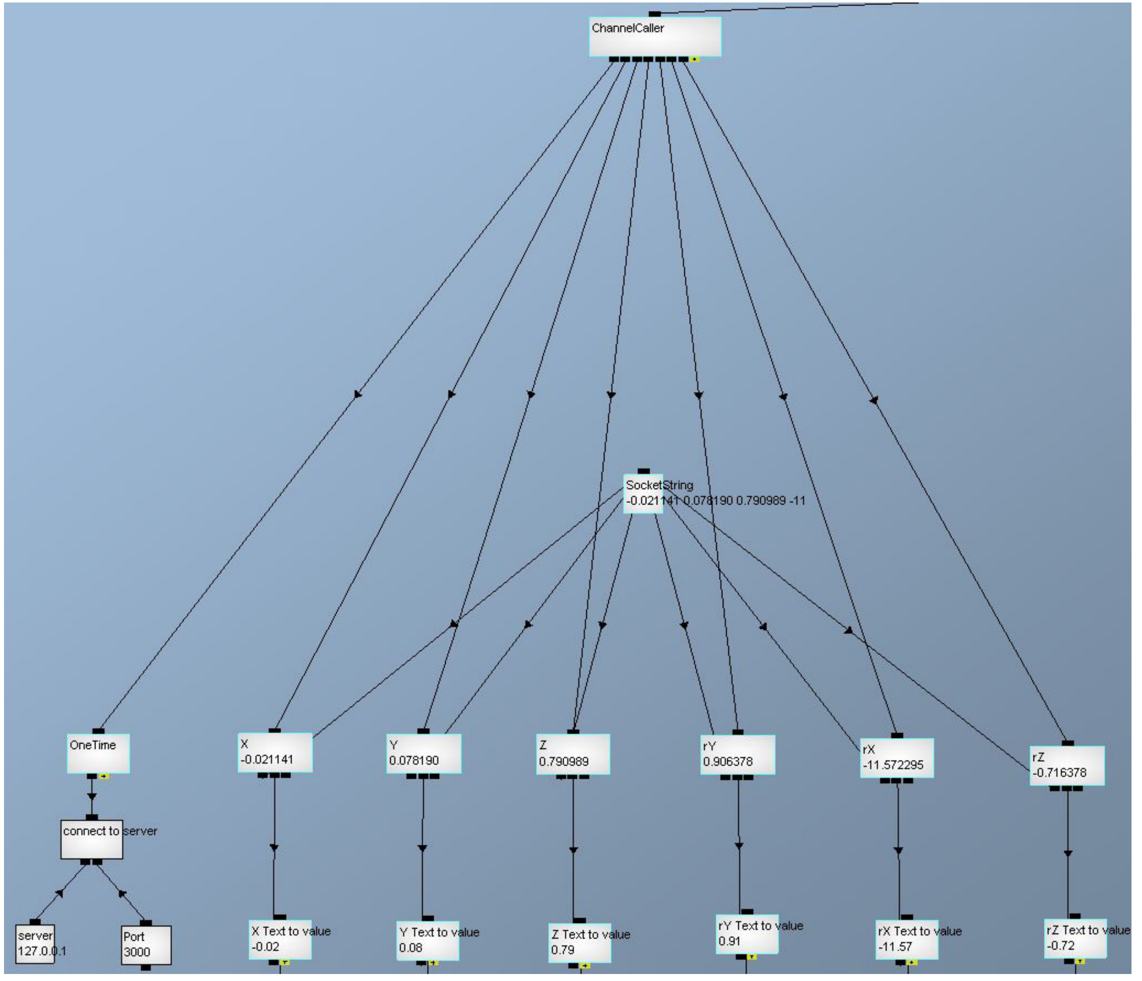


figure 6.5: Fetching the tracking data in Quest 3D

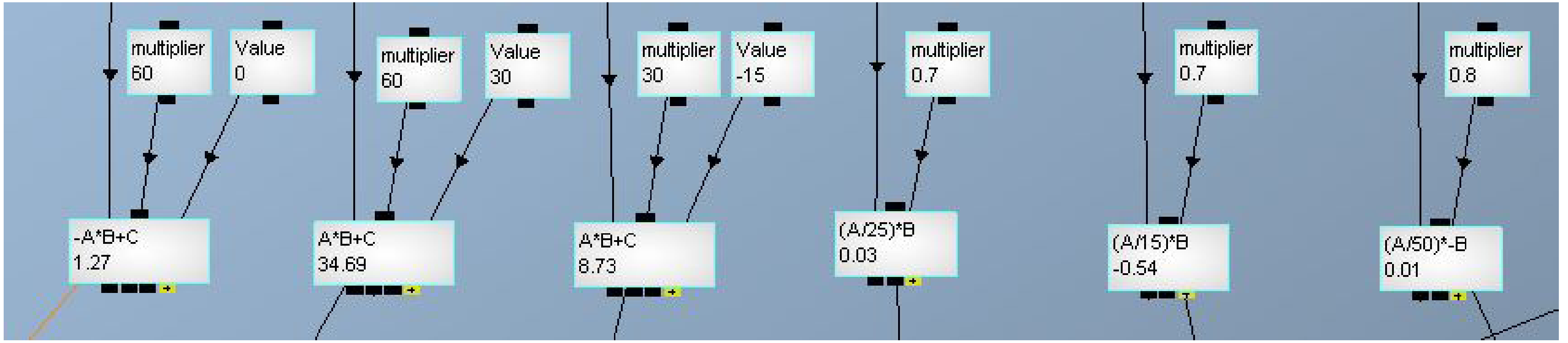


figure 6.6: Converting raw data to usable data

6.5.2 CONVERT RAW DATA TO USABLE DATA

The received tracking data will ultimately control the camera in the virtual environment. To make the raw data usable to control a camera it has to be converted to data that lies within the usable range in Quest 3D. The vectors are multiplied and an offset is added to give the camera the correct starting position. The most efficient way to do this was by trial and error, while making smaller adjustments at each iterative step.

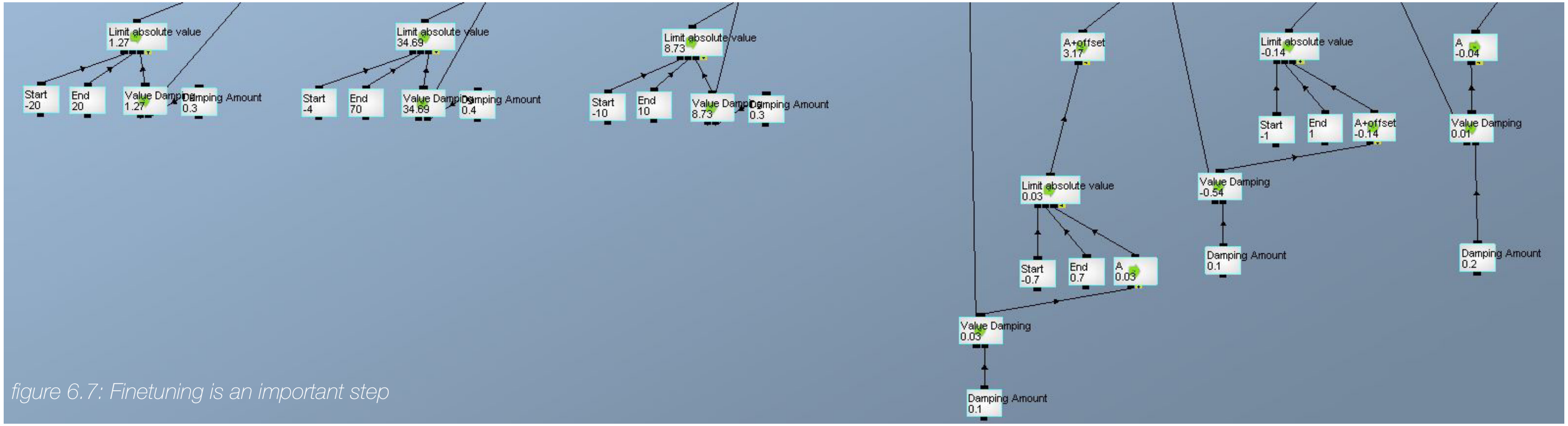


figure 6.7: Finetuning is an important step

6.5.3 FINE-TUNING MOTION

Although the values are usable in terms of range they do not deliver a positive experience yet. The camera keeps stuttering which undermines the overall effect of a perspective corrected display. Also, each unconscious head movement will result in a motion on screen, which did not feel natural.

Damping is added to prevent stutter and to make the motion on screen more fluid. This eliminated most of the stuttering and prevented movements on screen, while the head was in a stationary position. However, making the damping factor too large introduces lag in the motion of the camera which will undermine the overall effect.

6.5.4 POSITIONING CAMERA

After fine-tuning the values they are bundled in matrices called the motion vector and the rotation vector. Together with a preconfigured size-vector they form the camera-matrix. This matrix serves as input for the camera-component in which more camera-parameters can be configured.

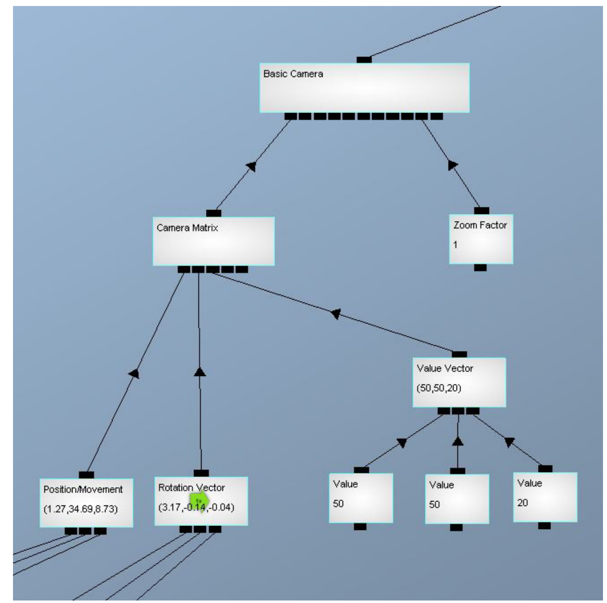


figure 6.8: Positioning the camera. In the lower left corner are the camera matrices.

6.5.5 CREATING VE AND VR-OBJECTS

All the above would go unnoticed if it were not for the virtual environment and the virtual object within it. This basic production is focused mainly on getting the tracking component to work, thus a simple environment is used that was readily available as an example in Quest 3D. Some random objects, a chair and a floating cube, are placed within the environment and a basic directional light is used to create shadows.

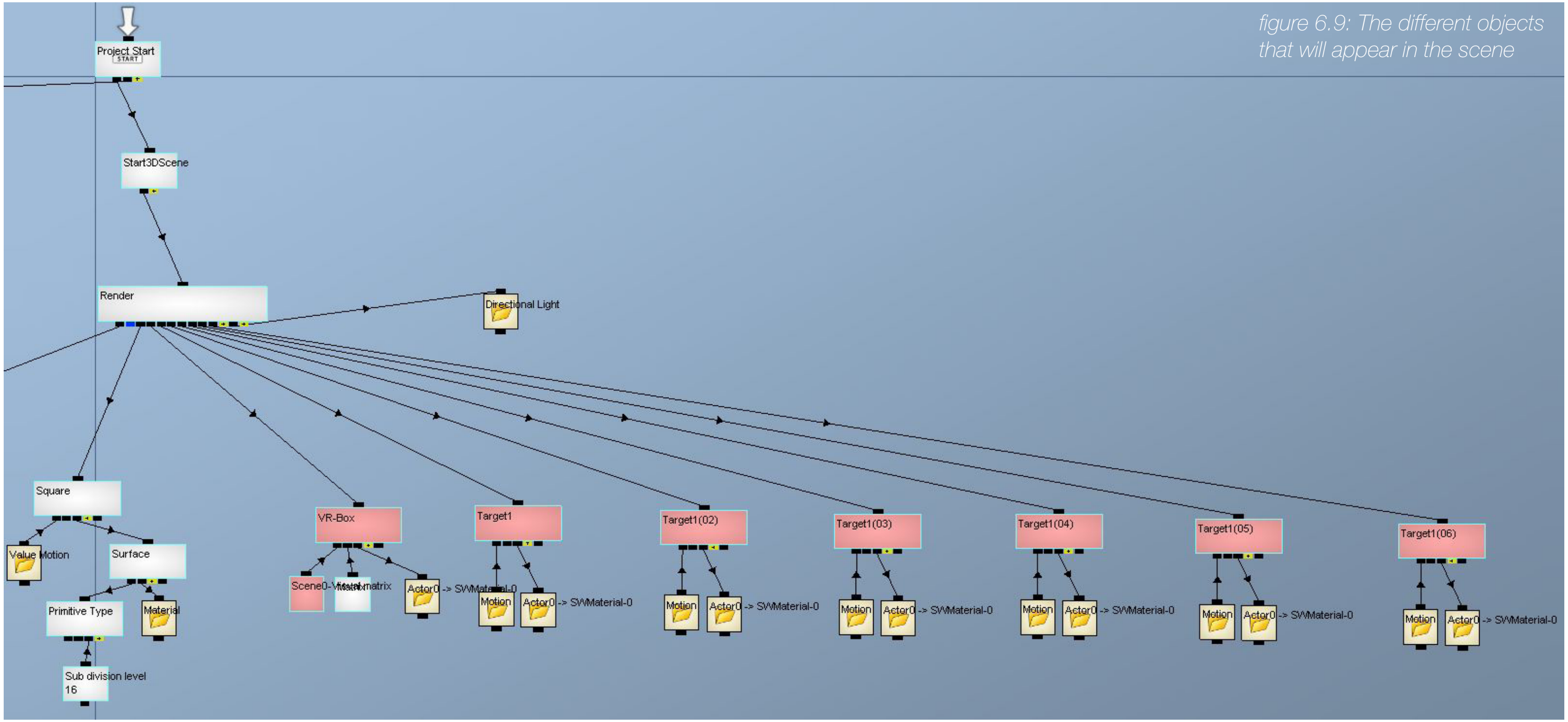


figure 6.9: The different objects that will appear in the scene

6.6 Refinement

6.6.1 EVALUATION

Building the tracking component in Quest 3D took considerable time, partly because of limited experience with the application, and partly because a lot of trial and error was involved. The 'quest' has been very iterative and new directions have been tried until a satisfactory solution was found.

The proof of principle technically works, but more careful evaluation showed it needs further improvement. Head movement seems to respond fairly well, but should feel more natural. The virtual environment is also to blame for this. The amount of depth cues present in this basic environment are limited and it is not very engaging. There are no restrictions to the camera movement, which makes it difficult to use as perspective corrected display. The user has to actively keep looking at an object when a head movement is made. Because the fluidity of the movement is still somewhat quirky, it is easy to lose the object out of sight. On the other hand, navigating the entire environment by just using head movement was a very pleasant experience that increased immersion.

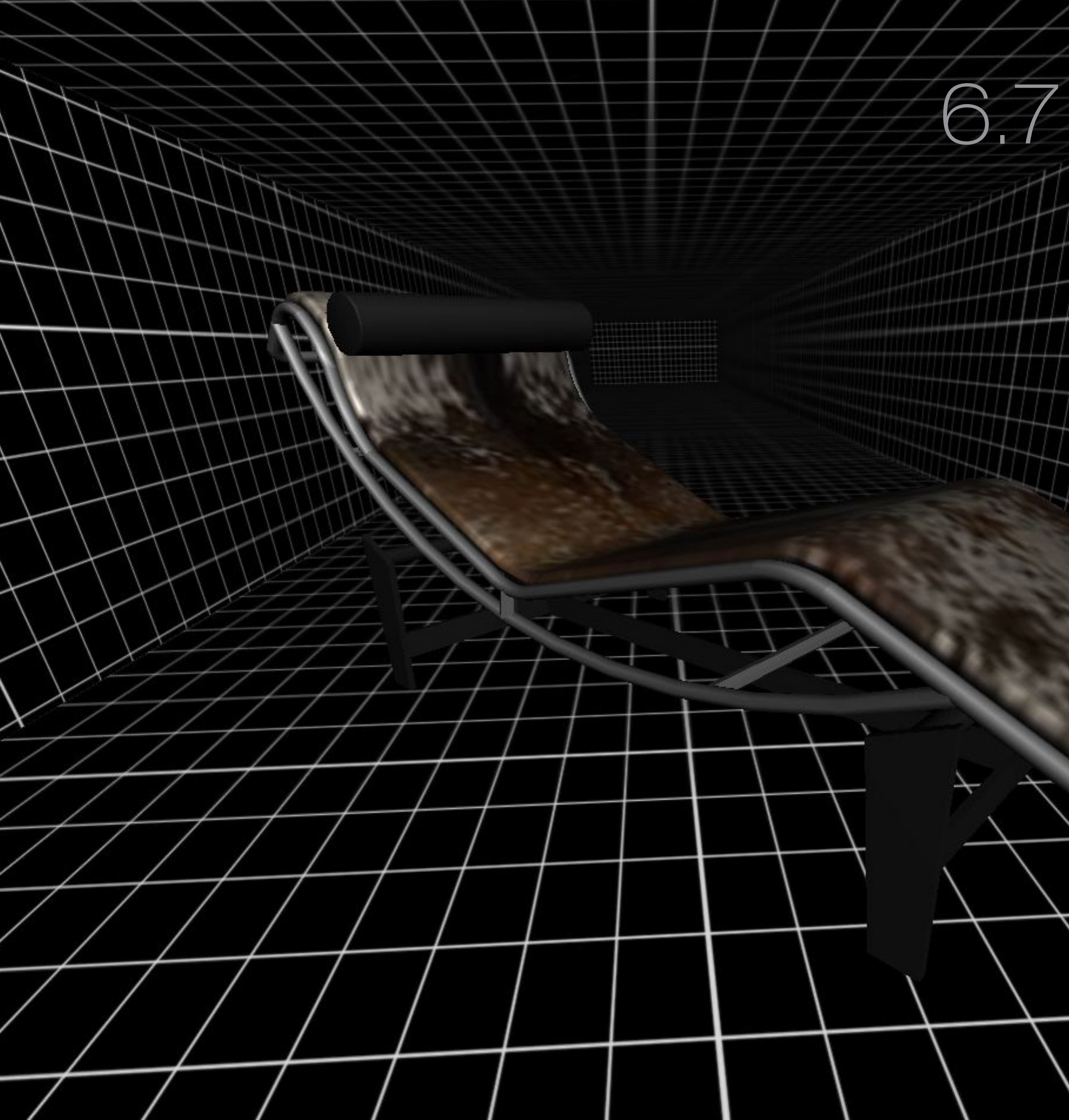
6.6.2 IMPROVED VERSION

The raw data has been modified based on the insights of the evaluation. Both the sensitivity of the motion and the damping factor has been adjusted and camera movement feels much more natural now. Also, two alternatives for camera movement are now possible; one with unrestricted camera movement to navigate an entire virtual environment and one with limited camera movement to support a perspective corrected view.

Another dramatic improvement is the use of a new virtual environment completely aimed at increasing the effect of a perspective corrected display. A long VR-box is created with a grid texture. Also the VR-objects intended for testing vary more in depth and will easily occlude each other. The environment might be a bit clinical, but it works well for its intended purpose.

There is still some stuttering visible, which is likely due to the use of an UDP-stream and the many transitions until the data generated in FaceApi reaches Quest 3D.

figure 6.10: The Corbusier LC04 chaise longue inside the VR-box



6.7 Conclusion

The solution route presented to implement a perspective corrected display in the VR-lab is pragmatic and works. However, there is a lot of room for improvement which mostly has to do with software licenses and compatibility between software.

Correcting the perspective to the position of a user's head is not enough to create a convincing experience. It is very important to make sure enough depth cues are present. Especially the use of a VR-box yielded a significant increase in depth perception.

The developed tracking component in Quest 3D is the core of the perspective corrected display. However, evaluation showed the tracking component could also successfully be used as an input method. By adding some more components in Quest 3D a user would be able to navigate a virtual environment (e.g. architectural walktrough) without the use of legacy input devices like a mouse or a keyboard.

7 Conclusions and recommendations

figure 7.1: A cockpit simulation in Quest 3D using headtracking

7.1 Conclusions

1. The key to designing a system which supports realistic control of virtual environments is to map all variables that influence the system. The performance of the tracking technology, that forms the foundation on which the system is build, is very dependent on the environment it is used in. A certain technology can perform well on all fronts in one environment, but might be absolutely useless in another environment.

2. It is recommendable to use an assessment tool to maintain overview of all parameters and assess what solution route is best. The tool simplifies decision making by focusing on issues raised by variables, instead of relying on the knowledge a user has of virtual reality and tracking technologies.

3. A time-of-flight camera would theoretically deliver the best tracking performance in the virtual reality lab. Based on gained knowledge and the assessment tool, this tracking technology would be relatively invariant to the difficult conditions in the VR-lab. Also, it has very high accuracy.

4. Webcam based tracking is the most viable solution to use within the scope of this thesis. The hardware is cheap and ubiquitous and tracking software is available in a free, non-commercial version. Although it is not ideal to implement, the performance of the system is solid and convincing.

5. The overall effect of a perspective corrected display is dependent on the VR-application and the amount of depth cues present. (e.g. conversing lines, occlusion) The more apparent the motion parallax is, the better the depth perception of the application will be.

7.2 Recommendations

1. The assessment tool should be elaborated to support better decision making of novice users. Currently the tool only shows possible issues. Without basic prior knowledge of perspective corrected displays and tracking technologies it is not always apparent when performance suffers. Users should be presented with a short description of all variables that play a role. Also it should be explained what consequence an issue has for the system.

2. User tests should be executed to assess when a perspective corrected display is helpful. A perspective corrected display is a building block and not a complete application. That is why it should first be implemented in existing applications in the VR-lab to enable testing.

3. A "channel" of the tracking component should be made to ease implementation in other Quest 3D projects. This will support its use in the VR-lab. The channel could serve as a front-end of the tracking technology used.

4. To achieve better performance and robustness of the system, a time of flight camera should be used for tracking. Algorithms based on Viola & Jones' framework, which are tailored to the data of a time of flight camera, already exist and might be obtained.

A smaller step towards increasing the performance of the system could be to use the Microsoft Kinect, which went on sale during this thesis and is now available in the VR-lab. Open source drivers have recently been released making its use on a computer much easier.

8 References

6DOFstreamer (2011) Project page on Google Code, http://code.google.com/p/6dofstreamer/

Act-3D (2011) Quest3D product information, http://quest3d.com/product-info/

Ascension technology corp. (2009) Magnetic Fact Sheet and Video: DC vs. AC Tracking, Ascension whitepaper, 2009

Bits & Chips (2011) Belgische 3D-camera gaat concurrentie aan met Kinect, http://www.bits-chips.nl/nieuws/bekijk/artikel/belgische-3d-camera-gaat-concurrentie-aan-met-kinect.html

Böhme M, Haker M, Riemer K, Martinetz T, Barth E (2009) Face detection using a time-of-flight camera, Proceedings of the DAGM Workshop on Dynamic 3D Imaging, pp. 167–176, 2009

Brandt E (2006) Designing exploratory design games: a framework for participation in Participatory Design? Proceedings of ACM Conference on Participatory Design, pp. 57-66, 2006

Carlson W (2008) A Critical History of Computer Graphics and Animation, section 17: Virtual Reality, http://design.osu.edu/carlson/history/lesson17.html

Cruz-Neira, Carolina, Sandin D J, DeFanti T A (1993), Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, Computer Graphics, SIGGRAPH Annual Conference Proceedings, 1993

Cybertherapy (1998) VR: An overview, http://www.cybertherapy.info/pages/introduc.htm

Davies RC (2004) Adapting virtual reality for the participatory design of work environments, Computer Supported Cooperative Work 13, pp. 1–33, 2004

Djajadiningrat J P, Smets G J F, Overbeeke C J (1997) Cubby: A multiscreen movement parallax display for direct manual manipulation, Displays 17, pp. 191–197.

DPI Animation House (2009) FutureLand uitgerust met nieuwste technieken van DPI Animation House, Press release DPI, http://www.dpi.nl/nl/newsItems/view/67

Favalora G E, Dorval R K, Hall D M, Giovinco M G, Napoli J (2001) Volumetric three-dimensional display system with rasterization hardware, Proceedings of SPIE, Vol. 4297A, pp. 227-235, 2001

Fourward Technologies (2010) Dual-Purkinje-Image concept, http://www.fourward.com/dconcept.htm

Godbersen H (2011) UDPreader, http://godbersen.eu/0198e29e6b1026504/0198e29e6b1074614/index.htm

Haller M, Billinghurst M, Thomas B H (2007) Emerging technologies of augmented reality: interfaces and design, p. 44, 2007

How Stuff Works (2011) How Elumens Vision Station Works, http://computer.howstuffworks.com/elumens.htm

How Stuff Works (2011) How virtual reality works, http://electronics.howstuffworks.com/gadgets/other-gadgets/virtual-reality8.htm

LaViola Jr. J J (2009) The Wii Remote and You, CHI 2009 Course Notes, p. 216, 2009

Lee J C (2008) Hacking the Nintendo Wii Remote, IEEE Pervasive Computing 7, pp. 39-45, 2008

Li D (2006) Low-cost eye-tracking for human computer interaction, pp. 28-49

Li D, Parkhurst D (2006) Open-Source Software for Real-Time Visible-Spectrum Eye Tracking, Conference on Communication by Gaze Interaction (COGAIN), pp. 18–20, 2006

Lopez-Gulliver R, Yoshida S, Yano S, Inoue N (2008) gCubik: A cubic autostereoscopic display for multiuser interaction: grasp and group-share virtual images, Proceedings of ACM, SIGGRAPH Posters, 2008.

Luo X, Kenyon R, Kamper D, Sandin D, Defanti T (2007) The Effects of Scene Complexity, Stereovision, and Motion Parallax on Size Constancy in a Virtual Environment, Proceedings of IEEE Virtual Reality, pp. 59–66, 2007

Magic Vision Portal (2011) Project page on SourceForce, http://sourceforge.net/projects/magicvisionport/

Microsoft (2007) Microsoft Launches New Product Category: Surface Computing, Microsoft press release, http://www.microsoft.com/presspass/press/2007/may07/05-29mssurfacepr.mspx

Nongenre (2011) How Kinect senses depth http://nongenre.blogspot.com/2010/12/how-kinect-senses-depth.html

Oggier T, Lehmann M, Kaufmann R, Schweizer M, Richter M, Metzler P, Lang G, Lustenberger F, Blanc N (2004) An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger), SPIE Optical Design and Engineering, pp. 534-545, 2004

Perez A, Cordoba M L, Garcia A, Mendez R, Munoz M L, Pedraza J L, Sanchez F (2003) A precise eye-gaze detection and tracking system, 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, 2003

Prime Sense, Ltd. (2007) Method and system for object reconstruction, World intellectual Property Organisation (WIPO) patent number wo2007043036

Richardson D, Spivey M (2004) Eye-Tracking: Characteristics and methods, Encyclopedia of Biomaterials and Biomedical Engineering, 2004

RPC Photonics (2011) Diffractive optical elements, http://www.rpcphotonics.com/optical.asp

Seeing Machines (2011) FaceApi general FAQ, http://www.seeingmachines.com/product/faceapi/faq/

Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, Blake A (2011) Real-Time Human Pose Recognition in Parts from Single Depth Images, CVPR, 2011

Silicon Intelligence (2006) Viola and Jones' detector, http://www.siliconintelligence.com/people/binu/perception/node47.html

Stavness I, Lam B, Fels S (2010) pCubee: A Perspective-Corrected Handheld Cubic Display, Proceedings of CHI, pp. 1381-1390, 2010

Sutherland I E (1965) The ultimate display, Proceedings of IFIP Congress, pp. 506-508, 1965

Svanæs D, Seland G (2004) Putting the Users Center Stage: Role Playing and Low-fi Prototyping Enable End Users to Design Mobile Systems, Proceedings of the Conference on Human Factors in Computing Systems, pp. 479-486, 2004

Torrelle S, Dermaut E (2009) Masterproef: Actor recognition in digital video, pp. 26-28

UDP-TCP-bridge (2011) Project page on Google Code, http://code.google.com/p/udp-tcp-bridge/

VIEW (2003) Introduction into Virtual Reality, Information Society Technology program (IST), http://www.view.iao.fraunhofer.de/introduction.html

Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features, CVPR (1), pp. 511 – 518, 2001

Viola P, Jones M (2001) Robust Real Time Object Detection, IEEE ICCV Workshop Statistical and Computational Theories of Vision, 2001

VR-lab University of Twente (2011) Possibilities, http://www.VRlab.ctw.utwente.nl/

Weingarten J W, Gruener G, Siegwart R (2004) A state-of-the-art 3d sensor for robot navigation, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004

Wikipedia (2011) Application programming interface, http://en.wikipedia.org/wiki/Application_programming_interface

Wikipedia (2011) Astigmatism http://en.wikipedia.org/wiki/Astigmatism

Wikipedia (2011) Inertial navigation system, http://en.wikipedia.org/wiki/Inertial_guidance_system

Wikipedia (2011) Purkinje images, http://en.wikipedia.org/wiki/Purkinje_images

Wikipedia (2011) Sensorama, http://en.wikipedia.org/wiki/Sensorama

Wikipedia (2011) Stereo triangulation, http://en.wikipedia.org/wiki/Range_imaging#Stereo_triangulation

Wikipedia (2011) Time-of-flight camera, http://en.wikipedia.org/wiki/Time-of-flight_camera

Wikipedia (2011) Transmission control protocol, http://en.wikipedia.org/wiki/Transmission_Control_Protocol

Wikipedia (2011) Triangulation, http://en.wikipedia.org/wiki/Triangulation

Wikipedia (2011) User datagram protocol, http://en.wikipedia.org/wiki/User_Datagram_Protocol

Wikipedia (2011) Virtual reality, http://en.wikipedia.org/wiki/Virtual_reality

Woodman O (2007) An introduction to inertial navigation, Technical Report 696, University of Cambridge, 2007

Xsens Technology (2009) Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors, Xsens whitepaper, 2009

Annex A Plan of approach

ASSIGNMENT

The assignment is to design and make a perspective corrected display system for the Virtual Reality lab of the University of Twente. This system allows the user to see a 3D image on a 2D screen. The assignment will roughly consist of three phases:

In the first phase different perspective corrected display technologies will be researched as well as different technologies that can determine the user-position (e.g. head tracking). The different technologies will be mapped and, if possible, tested.

In the second phase a perspective corrected display system that addresses the needs and wishes of the VR-lab is designed. The software program 'Quest 3D' should be used to receive sensory input and to deliver output on any screen. Quest 3D is software that supports different types of input and is already used to generate and build virtual environments. The perspective corrected display system should also work on horizontal placed screens if that is possible.

The final phase is to develop a prototype device/program that can be used as a building block in the VR-lab.

DEFINITIONS

PERSPECTIVE CORRECTED DISPLAY

A display that shows an image that is corrected to the users perspective.

MOTION PARALLAX

Parallax is an apparent displacement or difference in the position of an object viewed along two different lines of sight. When stereopsis/binocular disparity is not available motion parallax can be achieved by moving ones head to gain different viewpoints.

HEAD TRACKING

A positioning system that tracks the position of the head or the eyes of a user.

ACTOR ANALYSIS

The actor is the Virtual Reality lab (T-Exchange) of the University of Twente. The Virtual reality lab started as a joint venture between Thales Netherlands and the University of Twente with the goal to mutually speed up innovations by exchanging knowledge.

The VR-lab is constantly being refined and upgraded, to keep up with the latest technologies. A holographic projector is one of the examples of new techniques that are used/tried. The design of a perspective corrected display is another building block that can help improve current services and/or expand the services of the VR-lab.

PROJECT FRAME

Virtual environments have become a commodity and they are increasingly realistic. The VR-lab features a variety of virtual environments that enables users to analyze and test products and/or situations. To get an immersive experience the virtual environments should be analogous to real life experiences. Currently, the virtual environments are controlled by user input, ranging from a conventional keyboard to multitouch interfaces.

Perspective corrected displays are used in combination with positioning systems and can offer a new building block to the VR-lab, which supports realistic control of virtual environments. A sub domain of positioning systems is head-tracking. Head-tracking calculates the position of the head or the eyes of the user relative to the screen on which an environment is displayed. This results in a life-like experience that is somewhat like looking through a window. When a user wants to see the left side of an object, one simply moves to the left and the environment adjusts to his or her viewpoint.

The big advantage of the use of head-tracking is that it can be used to generate three dimensional output on any conventional screen. This can be achieved by using the head-tracking input to create motion parallax. This is the apparent displacement or difference in the position of an object viewed along two different lines of sight (the user's eyes).

OBJECTIVE

The goal of the assignment is to design and make a perspective corrected display system for the Virtual Reality lab of the University of Twente.

The goal is useful because the system adds new functionality to the Virtual Reality lab. The system can be used to improve on services or add new ones.

The goal is executable, although the matter is relatively complex. Several alterations of components that are needed for the system to work are available. There is extensive information available about the different components (e.g. head tracking) on the internet.

STRATEGY

The research part will be an in-depth investigation to find possible options for the application of perspective corrected displays in the Virtual Reality lab. The research is in-depth because the design is directed to a very specific purpose. There is a large analysis phase that flows into a design phase. The research is qualitative oriented. Information is gained by both desk research as empiric research, because different perspective corrected displays are analyzed in both literature as in existing applications. Depending on the suitability of a technique it is decided if a concept will be created that meets the demands and wishes.

RESEARCH QUESTIONS

PHASE 1: RESEARCH

1. What is a perspective corrected display?
    a. Are there different perspective corrected display technologies?
    b. What are the advantages of perspective corrected displays?
    c. What are the disadvantages of perspective corrected displays?
    d. What are the applications of current perspective corrected displays?

2. What would be interesting use scenarios of 3-D images on 2-D screens?
    a. What do industrial designers think are interesting use scenarios?
    b. What are the benefits of a scenario in comparison to current scenarios?

3. What kinds of tracking technologies are available?
    a. How does a certain technology work?
    b. What are the advantages of a certain technology?
    c. What are the disadvantages of a certain technology?
    d. What are the uses of a certain technology?
    e. What are the findings from experimenting with a certain technology? (optional)

PHASE 2: DESIGN

4. Which system is best suited for the VR-lab?
    a. What are the demands the system has to comply to?
    b. What are the wishes the system could fulfill?
    c. Taking the demands and wishes into account, what are the possibilities for perspective corrected displays in the VR-lab? (e.g. surface table)

5. Which type of head-tracking fits the VR-lab best?
    a. What are the limitations head-tracking encounters in the VR-lab? (e.g. dark room, signal noise)
    b. Which tracking components are readily available? (e.g. webcam)

PHASE 3: IMPLEMENTATION

6. Which design elements should be executed?
    a. Is an element necessary to demonstrate the concept?
    b. Is an element executable?

7. Which steps need to be taken to execute the design?

8. Follow the aforementioned steps and create a prototype of the concept.

9. Testing the prototype.

MATERIAL

PHASE 1: RESEARCH

1. What is a perspective corrected display?
   Media > search engine (e.g. google)
   Literature > search engine (e.g. google scholar)

2. What would be interesting use scenarios of 3-D images on 2-D screens?
   a. What do industrial designers think that are interesting use scenarios?
   Persons > questioning experts, brainstorm with students
   b. What are the benefits of a scenario in comparison to current scenarios?
   Literature > search engine

3. What kinds of tracking technologies are available?
   Media > search engine (e.g. google)
   Literature > search engine (e.g. google scholar)

PHASE 2: DESIGN

4. Which system is best suited for the VR-lab?
   a. What are the demands the system has to comply to?
   b. What are the wishes the system could fulfill?
   Principal > questioning, documentation
   c. Taking the demands and wishes into account, what are the possibilities for perspective corrected displays in the VR-lab? (e.g. surface table)
   Reality > VR-lab analysis
   Demands and wishes > possible applications of techniques

5. Which type of head-tracking fits the VR-lab best?
    a. What are the limitations head-tracking encounters in the VR-lab? (e.g. dark room, signal noise)
    b. Which tracking components are readily available? (e.g. webcam)
    Media > search engine (e.g. google)
    Literature > search engine (e.g. google scholar)

PHASE 3: IMPLEMENTATION

6. Which design elements should be executed?
    a. Is an element necessary to demonstrate the concept?
    b. Is an element executable?
    Design elements > comparison > demands and wishes as assessment.

7. Which steps need to be taken to execute the design?
    design elements > elaboration

8. Follow the aforementioned steps and create a prototype of the concept.
    Elaboration

9. Testing the prototype.
    Elaboration

Annex B Brainstorm session

To findout what would be interesting uses for perspective corrected images a brainstorm session was held. Several industrial design students and some tech-savvies contributed to this brainstorm.

HANDHELD PERSPECTIVE CORRECTED DISPLAY

The user holds a display in his/her hands and a perspective corrected image is shown on the device. The inbuilt digital compass and accelerometer are used as input for the system.

A combination of active and passive motion parallax is used to create the pseudo 3d effect. A low tech system will only incorporate passive motion parallax, and thus the user has to move in the right position to see the image in the correct perspective. A more elaborate system could also incorporate active motion parallax and provide a better use experience. The product will most likely feature multi-touch input, which enables intuitive interaction with the virtual model. The mediums for this type of use scenario are tablet computers, smartphones, or personal media players. The main benefit of this use scenario in comparison to existing scenarios is that it is very easy to communicate concepts to principals.

E-READER

The abovementioned concept could also be used on e-readers to accompany text. Instead of emailing a PDF document with text and 2d images of a concept, an interactive 3D PDF document could be send that uses perspective corrected images of the concept. Of course it is also viable to present such a document on a stationary (desktop) screen.

BODYPEEK

There are several appliances for perspective corrected displays in surgery. Most interesting about the use of this technology during surgery is the ability to peek inside the body while using endoscopy. Instead of using manual input and create passive motion parallax, position tracking (e.g. head tracking) is used to look around in the body. When the surgeon wants to see the side of an organ, a head movement will do the trick.

Another appliance is remote surgery. Remote surgery (also known as telesurgery) is the ability for a doctor to perform surgery on a patient even though they are not physically in the same location. Again, the system is used as input to control cameras or an endoscope.

MEDICAL SCANS EVALUATION

This brainstorm concept already exists and is patented by Apple:

In an industry that relies heavily on charts and graphical information, head-tracking technology would be welcomed by physicians. It could be used to view X-rays, CAT scans, MRI scans, anatomical diagrams and more. For example, a neurologist could review an MRI scan of the brain by turning their head or using hand gestures.

In the area of Industrial Design, this technology could be used to evaluate large cad assemblies in a similar manner.

CAVED IN

An area in which perspective corrected images can provide added value is simulations. The user is CAVEd in. Conventional 2D screens or rear projection screens are used to show perspective corrected images. The shape of the cave depends on the type of simulation, but is most likely rectangular.

Instead of placing a car prototype with different instrumentation screens in front of a large projection screen like is currently done in the virtual reality lab, one could use a cave. If a person looks down, the instrumentation becomes visible. When a person looks side wards through the virtual car window the landscape changes. The experience is more immersive.

Perspective corrected displays can be incorporated in a variety of simulators. (Note: Are flight simulators already perspective corrected?)

CORRECTION RATIO

Using a 1:1 correction ratio is the most natural way of controlling a perspective corrected display. However, in some use scenarios it is unpractical to use a 1:1 ratio. When a user wants to evaluate a large CAD assembly, it could be a more pleasant experience if the CAD model would rotate faster than what the head movement would suggest.

Because a perspective corrected image that does not move according to the natural ratio, can diminish the pseudo 3d effect, it is advisable that the ratio can be adjusted to personal preferences.

Annex C FaceApi brochure

faceAPI

Download now at faceAPI.com

The Real-Time Face Tracking Toolkit for Developers and OEMs

Technical Specifications

faceAPI allows you to integrate Seeing Machines world-class face-tracking technology into your product or software application. Now available under both development and production licenses, faceAPI provides your development team with the ability to quickly incorporate ultra-reliable face-tracking, with no image-processing or computer vision expertise necessary.

faceAPI is the only comprehensive, commercially supported solution for developing products that leverage real-time face-tracking.

Features

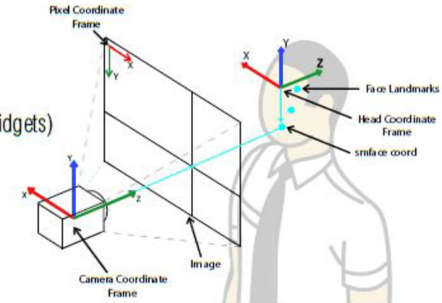
- Highly robust, real-time, 6 degree-of-freedom (3D) monocular face tracking
- Able to track up to +/- 90 degrees of head rotation and fast head movements
- Optional real-time tracking of lips and eyebrows
- Face-texture "mugshot" delivery upon tracking commencement
- Two types of head-trackers, trading CPU load for tracking capability.
- Requires only 8-bit greyscale video input. Algorithms do not rely on color and are therefore able to track faces in the dark using infra-red illumination
- Automatic tracking startup and immediate reacquisition when face is hidden then shown
- Facial feature detection and tracking – locates facial feature "landmarks" from a single image or real-time for video sequences
- Designed for all human faces (it doesn't matter what you look like). Robust to occlusion, fast movement, large head rotations, lighting changes, dimly lit rooms, facial deformation, skin color, beards, glasses etc
- Can track with as few as 40 pixels across the face (typically 2m from a VGA camera)
- Works with any webcam or video file
- Full control over all tracking parameters for purposes of software integration
- Low-level image input interface for custom camera integration
- "Offline" tracking algorithm for high-quality tracking of faces in movie files where real-time performance is not required
- Separate low-level "face-search" algorithm for finding multiple faces in images and short image sequences. This algorithm has three "depth" levels, with level 0 providing fast results to level 2 which accurately locates facial features and estimates head-pose in 3D.

Technical Details

At the heart of faceAPI lies a number of sophisticated multi-threaded tracking "engines". Each engine appears as a "black-box" with a minimal set of C functions that allow the developer to tune engine behaviour and performance.

faceAPI

The Real-Time Face Tracking Toolkit for Developers and OEMs



The API is provided as a closed-source library consisting of:

- C header files
- A DLL binary
- Concise HTML documentation
- Several sample applications (including a full game engine demo)
- A set of open-source object-oriented wrapper classes (C++ and Qt widgets)
- A tool for calibrating camera lenses (enabling precision tracking)
- A command-line tool for tracking faces in movie files
- Redistributable third-party APIs required for installation

faceAPI can track with a variety of video inputs:

- Webcams, or any WDM compatible camera (DirectShow driver)
- PointGrey Flea, Flea2 or Firefly MV cameras (recommended for high-performance applications)
- Movie files. Can read Windows avi or Apple Quicktime files.
- Low-level "shared-memory" image interface (allows for integration of custom video devices).

The tracking engine provides:

- Methods to "call back" your application whenever a new tracking measurement occurs
- A function that sets the current estimate of head-pose into a C structure (interpolates between samples)
- Head-pose measured in cartesian 3D coordinates relative to the camera, (X,Y,Z). Position is in meters and rotation is expressed in euler angles (rads).
- Positions of key facial locations (face-landmarks) including lip and eyebrow points, expressed in "face-coordinates" (inside the face, independent of head-pose)
- "Mugshot" face-texture is provided via a function callback that occurs when the face starts tracking. This texture is 256x256 resolution in RGB color format and is a synthetic orthographic projection of the front of the face. Face outline (texture-mask) points are provided with the texture.
- Head-pose measurements include a confidence weighting, from 0 to 1, allowing your application to determine when tracking quality is acceptable
- Latency of measurement is ~11ms (processing time) + exposure time. Eg, for a 30Hz camera, latency = ~44ms

Licensing

Non-Commercial, Development and Production license options now available.
Please contact us at bizdev@seeingmachines.com for more information on pricing, licensing and distribution.
Phone: +[61] 2 6103 4700
USA free call: 1-888-393-4293

Seeing Machines • US Tel: 1-888-393-4293 • www.SeeingMachines.com
Level 1 • 11 Lonsdale St • Braddon • ACT • 2601 • Australia
T: +61 2 6103 4700 • F: +61 2 6103 4701 • bizdev@seeingmachines.com

seeingmachines

www.jurriendijkstra.com